目　　录

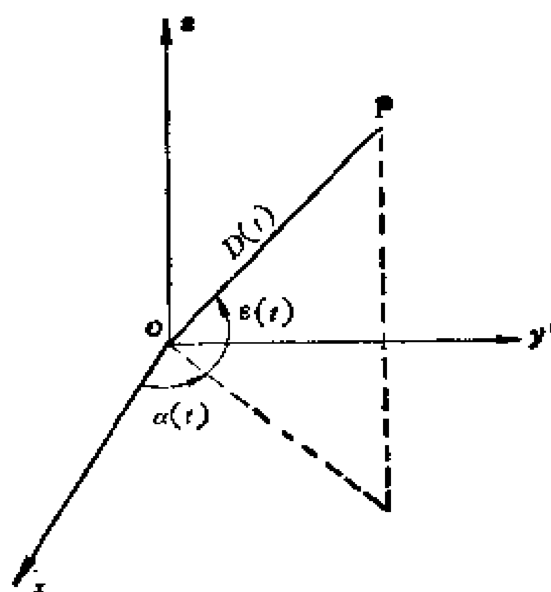
绪　　论

在工业、国防和科学技术的很多领域里，被考查的对象在其发展过程中，由于受到各种偶然因素的影响，往往表现出某种随机性，它们常常以一串随时间而变化的数据序列形式被人们记录到．这种序列通称为随机序列或时间序列，分析研究这些序列的数学方法就叫做时间序列分析方法，有时也称为时序分析．时序分析的目的是为了进行预报、滤波或控制等等．在时序分析的基础上，讨论时间序列的预报和随机漂移的控制等方法，是本书所称的时间序列的应用部分．

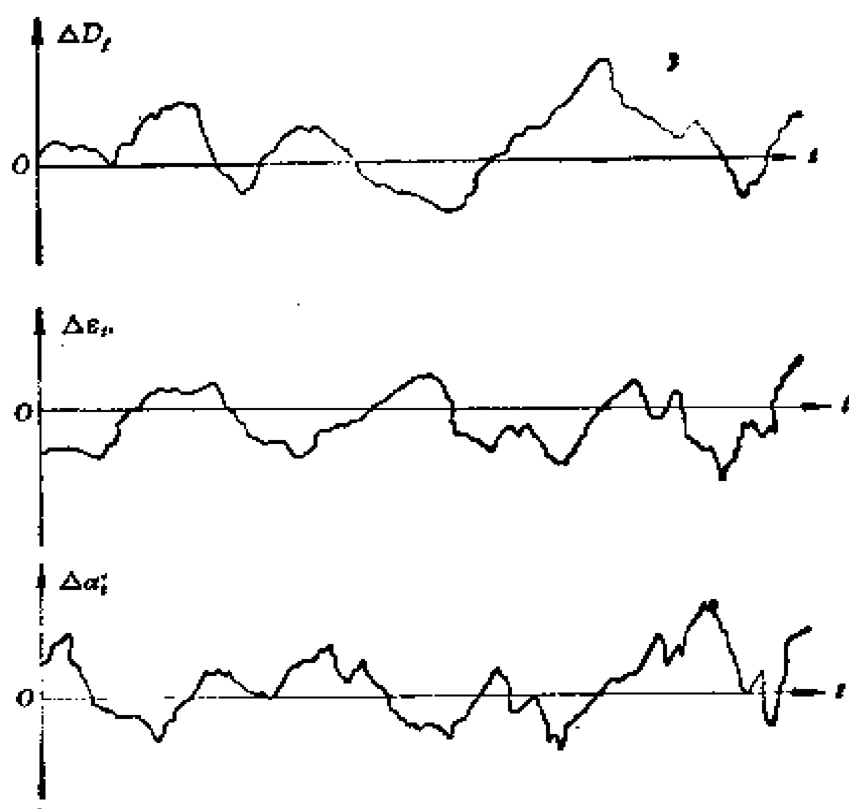
如何从具体的实际问题中提出时序分析问题？对这些问题如何分类？解决各类问题可以使用本书的哪些方法？我们将通过以下七个典型例子加以说明．

例 1. 跟踪雷达误差特性的鉴定问题．一部目标跟踪雷达的质量如何，最终要通过它跟踪目标时的测量误差来衡量．为了鉴定这种雷达，须用雷达跟踪飞行目标(如飞机)，记录其飞行轨道数据(见图一)，即目标与雷达的瞬时径向距离 $D(t)$，仰角 $\varepsilon(t)$ 和方位角 $\alpha(t)$．若取等间隔的采样，就得到三串数据序列 D_k，ε_k 和 $\alpha_k(k=1,2,\cdots)$，其中 $D_k=D(k\tau)$，$\varepsilon_k=\varepsilon(k\tau)$ 和 $\alpha_k=\alpha(k\tau)$ $(k=1,2,\cdots)$，τ 是采样间隔．与此同时，用更精确的观测设备(比如经纬仪照像设备)观测同一目标，并记录以上三个相应的坐标序列，其结果表示为 D_k^0，ε_k^0 和 $\alpha_k^0(k=1,2,\cdots)$．这些量相对于 D_k，ε_k 和 α_k 而言，可以近似为目标的真实轨道坐标数据．于是序列 $\Delta D_k=D_k-D_k^0$，$\Delta\varepsilon_k=\varepsilon_k-\varepsilon_k^0$ 和 $\Delta\alpha_k=\alpha_k-\alpha_k^0$ $(k=1,2,\cdots)$（见图二）就可以看做雷达对此目标跟踪时的测量误差序列．但是，即使在相同飞行条件下，重复试验所获得的误差序列也不会完全相同．那么用什么方法刻划这些误差序列的属

性呢？这就属于时序分析的课题.



图一 用雷达跟踪目标



图二 雷达跟踪飞行目标时的误差

众所周知，这些误差序列的样本均值与样本方差值分别由以下诸式计算:

$$\overline{\Delta D}=\frac{1}{N}\sum_{k=1}^{N}\Delta D_k,\quad \overline{\Delta\varepsilon}=\frac{1}{N}\sum_{k=1}^{N}\Delta\varepsilon_k,\quad \overline{\Delta\alpha}=\frac{1}{N}\sum_{k=1}^{N}\Delta\alpha_k \tag{1}$$

$$\begin{aligned}(\overline{\Delta D})_2&=\frac{1}{N}\sum_{k=1}^{N}(\Delta D_k-\overline{\Delta D})^2\\(\overline{\Delta\varepsilon})_2&=\frac{1}{N}\sum_{k=1}^{N}(\Delta\varepsilon_k-\overline{\Delta\varepsilon})^2\\(\overline{\Delta\alpha})_2&=\frac{1}{N}\sum_{k=1}^{N}(\Delta\alpha_k-\overline{\Delta\alpha})^2\end{aligned} \tag{2}$$

它们是描述序列 ΔD_k, $\Delta\varepsilon_k$, $\Delta\alpha_k$ 的重要数字指标. 但是，仅用这些指标还不能揭示出序列的更多更深刻的性质，也不能满足现代技术的需要. 我们将要介绍的时序分析方法是一种参数化的时域分析方法，它便于在数字计算机上实现. 所谓“时域分析”，是指通过数据序列在时间域上的各种平均值进行分析，比如用上述的样本均值、样本方差值以及更有用的样本协方差函数值等. 样本协方差函数值也是一种时域上的平均值，以 ΔD_k 序列为例，其定义为

$$\hat{\gamma}_j(\Delta D)=\frac{1}{N}\sum_{k=1}^{N-j}(\Delta D_k-\overline{\Delta D})(\Delta D_{k+j}-\overline{\Delta D})$$

$$j=0,\ 1,\ 2\cdots \tag{3}$$

至于“参数化”的含义，是说我们事先把序列的协方差函数

$$(\gamma_j(\Delta D),\ j=0,\ \pm1,\ \pm2,\ \cdots)$$

(其定义见第一章)，假定(或者说近似地)属于某一特殊函数类，即具有有理谱密度的平稳序列的协方差函数类(见第一、二章)，这一函数类中的每个函数被一组参数值唯一决定. 而根据附录 §1 可知，协方差函数 $\gamma_j(\Delta D)$ 属于这一函数类的充分必要条件是序列 $w_k=\Delta D_k-E(\Delta D_k)$ 满足以下的随机差分方程

$$w_t-\varphi_1w_{t-1}-\cdots-\varphi_pw_{t-p}=a_t-\theta_1a_{t-1}-\cdots-\theta_qa_{t-q}$$

(4)

其中 a_t 是均值为零的白噪声序列，即 $Ea_t = 0$, $Ea_ta_\tau = \delta_{t,\tau}\sigma_a^2$，而且由(4)式的系数确定的两个代数方程

$$1 - \varphi_1\omega - \varphi_2\omega^2 - \cdots - \varphi_p\omega^p = 0$$

$$1 - \theta_1\omega - \theta_2\omega^2 - \cdots - \theta_q\omega^q = 0 \tag{5}$$

的根全在单位圆外，即它们的模都大于1，而且两个方程没有公共根．我们把(4)式称为随机序列 w_t 的随机模型，或简称模型．这样一来，所谓对时间序列 ΔD_k 的分析，就转化成求定参数值 p, q 和 $\varphi_1, \varphi_2, \cdots \varphi_p$, $\theta_1, \theta_2, \cdots, \theta_q$, σ_a^2 的问题了．因此称这种时序分析为"参数化的时域分析"．第三、四两章就是论述这些参数的估计方法的．

根据大量的数据分析结果表明，在雷达跟踪同一空域的飞行目标时，其跟踪误差序列的模型可近似表为

$$w_t - \varphi_1 w_{t-1} - \varphi_2 w_{t-2} = a_t$$

无论是 ΔD_k, $\Delta \varepsilon_k$ 还是 $\Delta \alpha_k$，都可以用上述模型近似描述，只是各自的参数有所不同．由这种模型所决定的 w_t 的自协方差函数的形式包括以下的常见形式

$$\gamma_j = \sigma^2 e^{-\beta|j|} \cos \omega j, \quad j = 0, \pm 1, \pm 2 \cdots \tag{6}$$

通过这种时序分析，不仅能给出误差序列 ΔD_k, $\Delta \varepsilon_k$, $\Delta \alpha_k$ 的系统偏差量的估计，以及各自方差的估计，而且给出了序列的频谱特性的估计(见第一章)，它反映了雷达的跟踪性能．粗略地说，ω 和 β 的值越小，雷达对目标机动变化的反应越迟钝．至于分析这些误差特性的目的，对于研制雷达者来说，是为了了解产品的品质和改进方向，对于使用者来说，目的之一是为了决定能否接收产品．还有另外的目的，将在下一个例子中叙述．

上述分析内容不仅适用于雷达误差鉴定，对于许多连续工作的精密仪表的随机漂移量，比如惯性陀螺的随机漂移量等，同样可以进行类似的时序分析．

例 2. 用滤波方法解决测轨问题．在设计高射炮指挥仪或导弹制导时，都会遇到适时测轨问题．近年来常用 Kalman 滤波[20]方

法解决这类问题，而且通常是采用较简便的二阶多项式模型方程．这里我们仅叙述轨道的一个坐标分量 x_k 的滤波方程，其它坐标完全类似．设目标的瞬时速度为 $\dot{x}_k$，加速度为 $\ddot{x}_k$，取三维状态矢量为 $\boldsymbol{x}_k=(x_k, \dot{x}_k, \ddot{x}_k)^\tau$，$\boldsymbol{x}_k$ 的变化规律有时被近似地描述为

$$\boldsymbol{x}_k=\begin{pmatrix}1 & \tau & \frac{1}{2}\tau^2 \\ 0 & 1 & \tau \\ 0 & 0 & 1\end{pmatrix}\boldsymbol{x}_{k-1}+\begin{pmatrix}0\\0\\u_k\end{pmatrix} \tag{7}$$

其中 τ 是采样间隔(有时在(7)式中的三阶方阵的第三行第三列的元素 1 用 β 代替)．雷达对坐标分量的量测结果可表为

$$y_k=(1, 0, 0)\boldsymbol{x}_k+v_k \tag{8}$$

其中 u_k 和 v_k 是彼此独立的随机序列，u_k 称为系统噪声，v_k 为雷达的量测误差．为了利用 (7)，(8) 式和 Kalman 滤波方法适时估计状态矢量 $\boldsymbol{x}_k$，须要知道 u_k 和 v_k 各自的协方差函数或随机模型．其中对 v_k 的分析，完全属于例 1 的时序分析问题．由此可见，测定跟踪雷达的误差序列的模型参数，对使用者来说还有另一种用处，就是提供滤波方法中的部分参数值．当然还不只限于为滤波提供参数，也为控制设计等提供参数．

下面着重介绍如何测定 u_k 的随机模型．为此，也可以利用例 1 的飞行试验中所获得的数据资料 D_k^0，ε_k^0 和 α_k^0，或者转换成直角坐标分量 x_k，y_k 和 z_k，我们仅以坐标 x_k 为例说明相应的系统噪声 u_k 的分析方法．如例 1 那样，D_k^0，ε_k^0 和 α_k^0 被当做真实轨道数据，那么 x_k 也是真实轨道数据．由 (7) 式可知

$$x_k=x_{k-1}+\tau\dot{x}_{k-1}+\frac{1}{2}\tau^2\ddot{x}_{k-1}$$

$$\dot{x}_k=\dot{x}_{k-1}+\tau\ddot{x}_{k-1}$$

$$\ddot{x}_k=\ddot{x}_{k-1}+u_k$$

其中 $k=1, 2, \cdots$，由此不难看出，即使 u_k 可以近似地看做平稳序列，x_k 也不是平稳的．事实上，x_k 是飞行目标的坐标量，它的变化一般不会是平稳的．因此分析数据 x_k 不能和分析 v_k 的方

法完全相同．这时我们希望通过非平稳序列 x_k 的数据建立 u_k 的平稳模型，所以先要找出 x_k 和 u_k 的关系式．利用上面三个式子，分别计算 x_k 的二次和三次差分：

$$\nabla^2 x_k = (x_k - x_{k-1}) - (x_{k-1} - x_{k-2})$$

$$= \tau \dot{x}_{k-1} + \frac{1}{2}\tau^2 \ddot{x}_{k-1} - \tau \dot{x}_{k-2} - \frac{1}{2}\tau^2 \ddot{x}_{k-2}$$

$$= \tau(\dot{x}_{k-1} - \dot{x}_{k-2}) + \frac{1}{2}\tau^2 u_{k-1}$$

$$= \tau^2 \ddot{x}_{k-1} + \frac{1}{2}\tau^2 u_{k-1}$$

$$\nabla^3 x_k = \tau^2 \ddot{x}_{k-1} + \frac{1}{2}\tau^2 u_{k-1} - \tau^2 \ddot{x}_{k-2} - \frac{1}{2}\tau^2 u_{k-2}$$

$$= \frac{3}{2}\tau^2 u_{k-1} - \frac{1}{2}\tau^2 u_{k-2} \tag{9}$$

由此可见，$\nabla^3 x_k$ 可以近似地看做平稳序列．若将数据进行三次差分后，再按照例 1 的方法进行时序分析，并给出其模型参数的估计，然后利用 $\nabla^3 x_k$ 与 u_k 的关系式(9)便可以得到 u_k 的估计模型．假定用 w_k 代表 $\nabla^3 x_k$，其模型定为(4)式的形状，那么将(9)式代入(4)式可得

$$\left(\frac{3}{2}\tau^2 u_{t-1} - \frac{1}{2}\tau^2 u_{t-2}\right) - \varphi_1\left(\frac{3}{2}\tau^2 u_{t-2} - \frac{1}{2}\tau^2 u_{t-3}\right) - \cdots$$

$$- \varphi_p\left(\frac{3}{2}\tau^2 u_{t-p-1} - \frac{1}{2}\tau^2 u_{t-p-2}\right)$$

$$= a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}$$

令

$$\frac{2}{3\tau^2} a_t = b_{t-1}$$

则上述模型化为

$$u_t - \left(\varphi_1 + \frac{1}{3}\right)u_{t-1} - \left(\varphi_2 - \frac{1}{3}\varphi_1\right)u_{t-2} - \cdots$$

$$- \left(\varphi_p - \frac{1}{3}\varphi_{p-1}\right)u_{t-p} + \frac{1}{3}\varphi_p u_{t-p-1}$$

$$= b_t - \theta_1 b_{t-1} - \cdots - \theta_q b_{t-q}$$

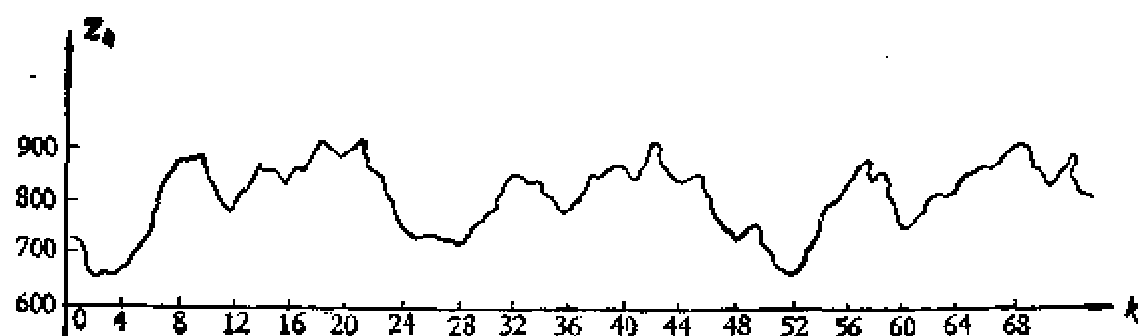
这就是 u_t 的随机模型. 若由上式的系数所确定的相应于(5)式的前一代数方程以 1 为根，则应考虑 x_t 四阶差分（详见第二章）.
通过这一例子的叙述，我们得到的另一启发是，时序分析方法可以推广应用于像 x_k 这样的非平稳序列，即 x_k 的若干次差分后的序列属于我们所讨论的平稳序列.

在应用滤波和控制的领域中，几乎都会遇到上述类似的问题.
近十几年来的时序分析方法随着现代滤波和控制的广泛应用而得到了新的发展. 本书所介绍的方法就是最近应用较多的方法.

例 3. 随机震动模拟问题. 很多精密装置在运输过程中因受震动而影响其寿命或质量. 为要了解某种装置因受这种震动所造成的损耗情况，常常需要进行大量的试验. 当然，可以把该装置放在火车或其他运输工具上，通过长时间运行后考查其损耗程度. 科学的方法是在室内设计一种震动台，将此装置放在台上，使台体模拟运输中的震动，以代替运输试验. 使震动台模拟运输震动的方法有很多种，过去常用多次谐波震动来模拟，但这种方法与真实情况相差较大. 现在多采用随机震动方法，也就是使震动台按照噪声发生器产生的随机序列进行震动. 于是自然提出这样的问题：用怎样的随机序列值驱动震动台，才能更逼真地模拟运输震动？解决这类问题便可用时序分析方法. 其大致步骤如下，先将震动敏感元件和记录设备安放在运输工具上，使其在某个路段上行驶并记录下震动变化量，再通过离散采样而得到一串时间序列值. 如例 1 所述，可以对此序列值进行时序分析，确定它的模型，然后用噪声发生器(或用计算机)产生出同类型的随机序列，使震动台按照这样的序列值进行震动. 这样模拟的效果显然更接近真实情况.

在以上三类应用中，进行时序分析工作之后，如何完成进一步的任务，如最佳滤波器和模拟装置的设计等等，这些都不属于本书的范围. 下面的几个例子中提出的问题归于另外几种类型的时序分析的应用，包括预报、识别、控制或检测等. 这些都在本书所讨论的范围之内.

例 4. 电力负荷预报. 以 z_k 表示第 k 小时内某城市平均用电负荷量,开始时刻取 $k=1$, 以后 k 递增下去. 序列 $z_k(k=1,2,\cdots)$便是具有一定随机性的时间序列(见图三).



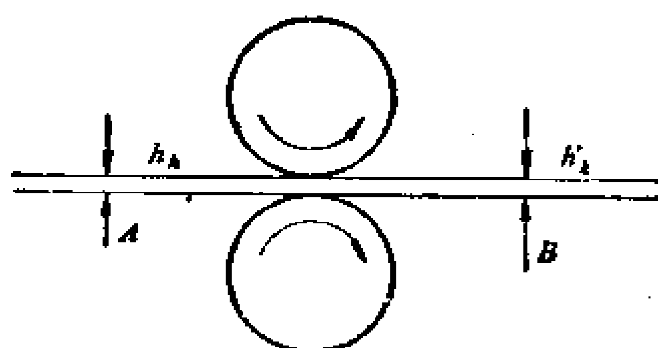
图三　某地各小时平均用电负荷曲线

为了合理分配供电,须要根据现在和过去时刻用电情况,估计未来时刻用电的多少. 比如现在时刻为第 n 小时,那么,就要利用数值 $z_1,z_2,z_3\cdots,z_n$ 预报 $z_{n+1},z_{n+2},\cdots$等值. 这就是电力负荷预报问题.从直观上看,每天同一小时(如上午八点)的用电量似乎是差不多的,也就是说,序列 z_k 近似为周期变化, 其周期长度为 24. 因此,可以用三角函数回归的方法进行预报. 对这类可称为"季节性模型"的时间序列, 本书将采用另一种方法处理, 即用类似于处理例 2 中 x_k 的办法, 对 z_k 的一种特殊的差分序列进行时序分析, 确定它的随机模型(详见第六章),然后用最小方差估计(见第一章和第七章)进行预报. 这样效果比较好.

解决各种预报问题, 是时间序列最广泛的一类应用. 比如在气象、水文、天文、交通及海洋研究等部门都可应用,还有人进行过预报病虫害的尝试.

例 5. 轧钢过程的系统识别问题. 在轧钢工艺中,钢板通过轧辊后, 被轧制成更薄的钢板(见图四). 控制轧制厚度是通过调整辊缝实现的. 所谓辊缝,是指上下两个轧辊之间的空转时的距离. 由于固定两个轧辊的相对位置的设备有一定的弹性, 所以在钢板输送到轧辊时,一方面轧辊滚动带动钢板前进,同时由于挤压而使

真实辊缝变大，因而轧制后的钢板厚度也就比辊缝略大些．在输入钢板的厚度和硬度为理想均匀情况下，可以按照轧制后的厚度要求，计算出合适的辊缝，并将它固定下来．这就是单机架的简单轧制过程．由于实际的钢板厚度和硬度是在某一均值上下起伏变化的，所以在固定辊缝轧制时，轧辊的相对位置总在变化，由此轧出的钢板的厚度也在不断变化．为了提高轧制精度，近代采用自动化调节辊缝的技术．为此，如图四所示，在 A，B 两处安装测厚



图四　单机架轧钢示意图

装置，根据输入和输出的钢板的厚度 h_k 和 h'_k，适时地确定辊缝的调节控制量 x_k 以使 h'_k 取得尽量小的起伏值．这时我们就要用到系统和系统识别的概念(见第八章)．通常是把轧制过程当做有输入输出的定常线性系统：设 h_k 的起伏量是 $x_k = h_k - h$，h'_k 的起伏量是 $y_k = h'_k - h'$，当辊缝固定时(即取调节量 $x_k = 0$)，x_k 与 y_k 满足系统方程

$$
\begin{aligned}
& y_t - \zeta_1 y_{t-1} - \zeta_2 y_{t-2} - \cdots - \zeta_r y_{t-r} \\
& = \omega_0 x_{t-b} - \omega_1 x_{t-b-1} - \cdots - \omega_s x_{t-b-s}
\end{aligned}
\qquad (10)
$$

并且假定 x_t 或它的某次差分是有有理谱的平稳序列．这样，就可以采用时序分析的技巧，根据实测数据值 $(h_1, h'_1)(h_2, h'_2)\cdots(h_N, h'_N)$ 来估计(10)式中的各个参数．详细讨论见第八章．在有了模型(10)式以后，如何分析轧钢的自动控制调节问题，将在下例中阐述．与本例相类似的实际问题在工业自动化方面是很多的．

例 6. 轧钢自动化的最佳控制问题．继续考虑例 5 的轧钢自动化设计问题，为要设计自动调节辊缝的方案，不仅要识别系统

(10) 的参数，还要把辊缝调节量 x_k 与输出钢板厚度 h'_k 的关系也按 (10) 式的线性系统加以描述，并且识别它的诸参数．比如这个系统被确定为

$$
\begin{aligned}
& y_t^* - \xi_1 y_{t-1}^* - \xi_2 y_{t-2}^* - \cdots - \xi_l y_{t-l}^* \\
& = \eta_0 x_{t-f} + \eta_1 x_{t-f-1} + \cdots + \eta_j x_{t-f-j}
\end{aligned}
\tag{11}
$$

其中 y_t^* 表示在输入钢板的厚度和硬度为均匀的条件下，纯粹由于 x_t 的改变而引起的输出钢板的厚度起伏量，或者称为随机漂移量．假定(10)式和(11)式已分别建立起来，在实际轧制过程中，又假定输出钢板的厚度漂移量是来自三种独立影响的叠加，即由于 h_k 的随机起伏，控制量 x_k 的变化，以及钢板硬度变化三种因素造成的．此外，还假定 h_k 和 h'_k 是可以适时记录的量，x_k 是可以适时改变的量，而输入钢板的硬度虽不能直接适时量测，但它的起伏量可以用例 1 的办法进行模型识别．那么，所谓轧钢最佳自动控制问题，就是如何根据 $h_k, h_{k-1}, \cdots$ 和 $h'_k, h'_{k-1}, \cdots$ 适时确定控制调节量 x_k，以使输出量 h'_k 具有尽量小的起伏．解决这类问题的数学方法将在第九章介绍．我们选用这个例子，主要是为了阐述本书所讨论的随机控制的物理背景，以及如何将具体的实际问题转化成数学问题．真实的轧钢自动化过程往往是非常复杂的，比如要用到多台机架联轧，所以，要灵活地使用本书的方法，或者使用第十章的多维序列方法才可能付之实践．

在工业自动化领域中，尤其是化工生产自动化中，类似于轧钢自动控制的问题很多，一般都是把问题化成这里所描述的数学问题后再加以解决．

例 7. 路基检测问题．设从甲地到乙地有一段铁路，在火车上装一套敏感震动的记录设备，以记录火车在这段铁路上行驶中的震动情况．火车每在这段路上行驶一次，便可以得到一段随机序列值 $w_1, w_2, \cdots, w_N$（比如 w_k 代表震动幅值），根据这 N 个数值，类似于例 1，可以用本书第三、四章的方法识别出 w_t 的随机模型，即估计出(4)式中的诸参数值．火车从甲地到乙地多次往返行驶，便可获得大批的数据资料 w_k，从而可以多次地估计出它的模型参数．

根据大量正常行驶的记载，可进一步对这些参数估计值进行统计分析，比如求它们的多维分布、均值和方差等. 以此为依据，就能对以后的各次行驶记录进行鉴别，若在某一次运行中，所得到的模型参数估计值落在正常范围以外，则认为路基须要进行检修. 详细地判别方法可以从第五、六章的内容中找到，这就是检测 w_t 的随机模型是否发生变异的问题. 在复杂的系统装置中，或者在可靠性要求很高的设备中，检测问题是经常遇到的. 本书所介绍的是解决这类问题的方法之一.

以上我们通过一些具体例子概述了时序分析的应用背景. 当然，实际应用比我们所叙述的要丰富得多. 此外，还必须指出，实际问题往往是相当复杂的，而本书的方法总有它的局限性. 因此，为要解决好生产或科研项目中的具体问题，必须对客观事物作深入了解，然后，在此基础上灵活地使用时序分析方法，甚至要综合使用多种有关的数学方法，才能取得有实际意义的效果.

第一章　随机序列

这一章是本书的预备知识．我们借助于实例，用较通俗的语言引入平稳随机序列的概念．然后介绍平稳随机序列的描述方法，并且对于平稳序列中的"频谱分析方法"、"相关分析方法"及"参数化方法"之间的关系给予简要的说明．这些内容是本章的重点．此外，还将介绍两种常用的估计方法，这也是为以后几章做准备的．

§1. 过程的直观分类

客观事物的发展过程总可用数量加以表现，在这些数量表现中，有的呈现为随机时间序列的形式，有的可以转化成随机时间序列的形式．下面，我们通过几个实例来说明各种数量过程与随机时间序列的联系．

例 1. 地球与太阳之间的距离，这是一个随时间而变化的量众所周知，它又是一个周期变化的量．

例 2. 在某一海域里的一个固定点处的海水温度，它也是随时间而变化的量．但是，它不象例 1 中的量有严格的周期性变化规律．

例 3. 地球每自转一周的时间，都知道，是一个不变的量，即每自转一周都是 24 小时．

例 4. 某城市每小时平均用电负荷量，在不同的小时有不同的值．它是一串离散的变化量．

对于以上四种量，从直观上可以进行以下分类：

象例 1 这样的量，有完全确定的变化规律，可以用数学方法准确地描述出来，我们称之为非随机过程，又由于它随连续时间而变化，所以叫做非随机的连续时间过程．

例 3 与例 1 相类似，也是一种非随机的过程，不过它只取常数值24小时. 此外，它的取值是按一天一天计算的，不是随连续时间变化的，所以称它为非随机的离散时间过程.

像例 2 这样的量，虽然可以记录下以往各时刻的海水温度数据，但是不能用数学方法准确地描述这个量的过去和未来的取值，这是因为海水温度的变化受到很多偶然因素的影响，它的变化也就呈现出某种程度的不确定性. 我们把这样的数量过程称为随机过程.

例 4 与例 2 相类似，用电负荷量也是随机变化量. 但是，它又不同于例 2，而是按每小时计算的，所以称之为随机序列.

以上这些关于自然界各种变化过程的分类，是人们对于客观世界认识的一种概括. 应当指出，不能把这些分类看成是僵死的、一成不变的东西，它们相对立而存在，在一定条件又可以互相转化. 在利用这些分类概念去分析认识实际对象时，也不要机械地套用一种概念或一种方法，而应当灵活地使用各种认识手段，以求更深入更准确地反映客观世界，从而达到更好地改造客观世界的目的. 比如，由于大量应用电子计算技术，有时须要将连续时间过程转化为离散时间序列，最常用的转化手段是对原过程进行等间隔采样，采样后的过程就是序列形式，它在一定程度上反映了原过程的各种属性. 反之，在某些特定的条件下，也可将序列转化成连续时间过程. 再如例 3 中所说的地球自转时间，从前，由于当时的生产和技术水平的局限，人们能够认识到地球有自转运动，而且自转一周的时间相等，这已经算是科学的认识了. 但是，在现在的技术条件下，当人们计算的时间的精确度可以达到毫秒以下的数量级时，结果发现地球自转一周并非用严格相同的时间，而是在24小时上下有几毫秒的起伏量. 这样，本来认为是不足道的取常数值的非随机序列，现在却必须用变化的量来表示了 如果仍用非随机序列，则为了反应上述的微小变化量的规律，须要很复杂的微分方程描述，而且由于这种方法的近似性和实际计算上的复杂性，常常得不到令人满意的结果. 事实上，影响地球自转稳定性的干扰

因素，除了一些靠近地球的大星球的引力外，还有数不尽的微小的干扰量，把这些当作偶然因素考虑不是没有道理的. 所以，已经有人试图用随机序列来描写例 3 中的过程量. 由此可见，对于上述从实际中概括出来的随机与非随机，连续时间与离散时间等概念，以及本书讨论的其他概念与方法，要灵活地使用它们，还要在实践中有所创新.

§2. 随机序列的数学描述

为了介绍本节的内容，我们先从非随机过程的描述说起. 比如，一个闭合电路上的电压是一非随机的连续时间过程；如果电源是单频率的交流电，描述这一过程只要用一个正弦函数

$$A\sin(\omega t+\varphi_0)$$

就够了，其中 A，ω 和 φ_0 是依电源电路而决定的参数. 再比如 §1 中的例 1，如果忽略对地球公转的微小扰动，根据天体力学列出公转轨道所满足的微分方程，它的解（为一椭圆）就是这一过程的描述. 又如§1中的例 3，当把地球自转时间近似看为常数时，这个非随机序列只取同一常数值. 总之，要掌握非随机过程的变化规律，一般归结为选用适当的函数，确定各种参数，或列出物理方程求其解，这样便可以（相对一定的实际需要或认识能力而言）准确地描述这些过程. 现在再来看 §1 的例 2 和例 4，虽然海水温度和用电负荷量都具有明显物理意义的周期性，但是它们不能用周期函数很准确地刻划. 从直观上容易想象，用普通的微分方程也不能精确地描述这些量. 这就是随机过程与非随机过程的本质差别. 对于随机过程，必须用不同的方法加以描述和研究.

在讨论描述随机过程的方法之前，我们先指出以下两点. 其一，随机过程表面上虽然杂乱无章（如绪论中图二和图三），但是，它既然是客观事物和数量表征，必然有其内在的规律. 其二，为了掌握和利用这些随机过程所表现出来的规律，需要一定的数学工具，这就是随机过程理论. 为了加深读者在这方面的认识，我们特列举六个按照不同格式设计的伪随机数列作为例子，图 1.2.1 绘出

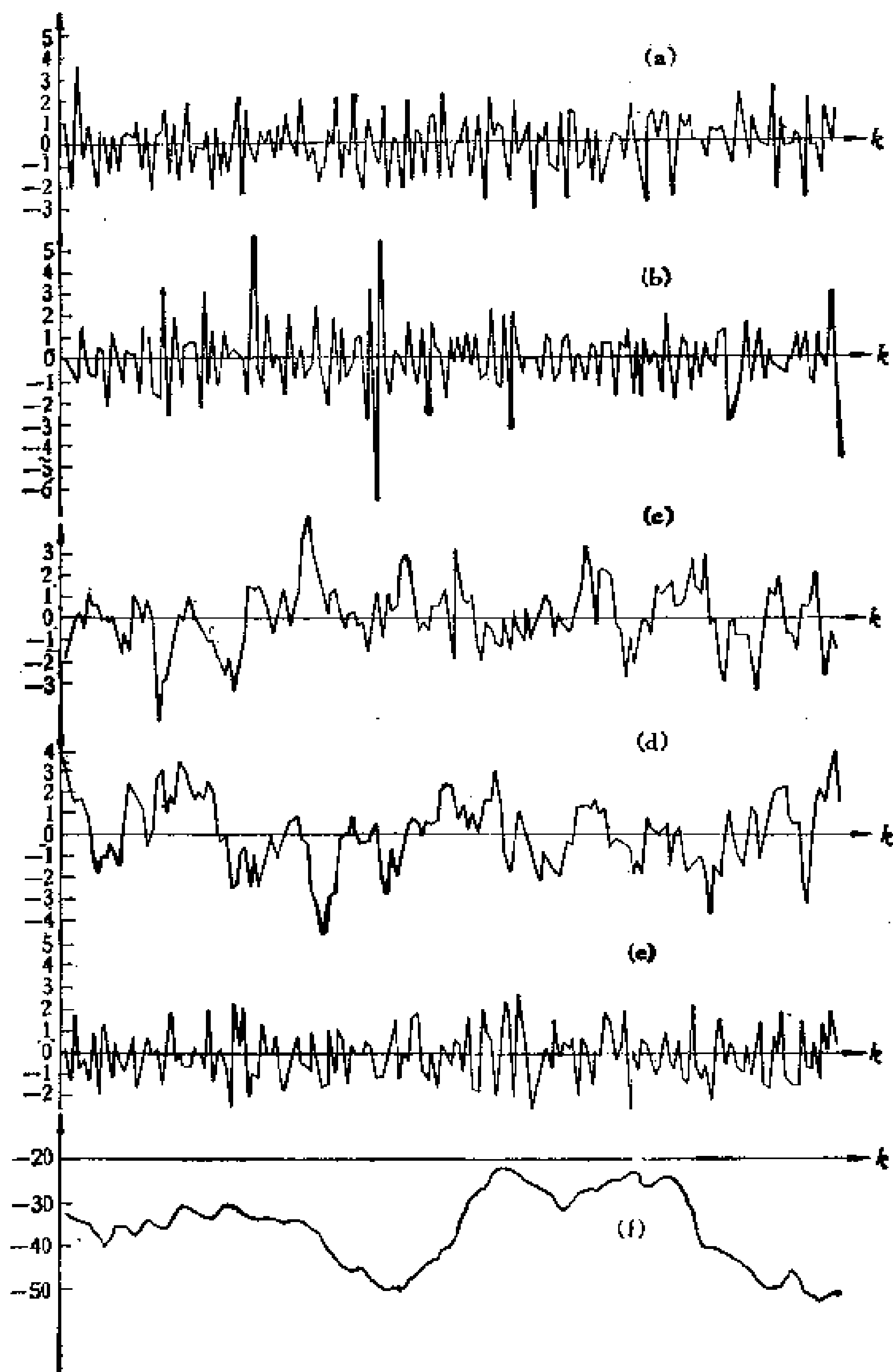


图 1.2.1

了它们的数值结果(用折线连接). 试问: 这些不同的数列反映了它们之间有哪些不同的属性呢? 又怎么掌握这些属性呢? 或者说,根据这些数列推测出它们是依怎样的设计格式计算出来的呢? 从图 1.2.1 看来,它们之中有的差别较大,有的似乎看不出什么差别. 但是,若用这六个数列分别计算出六个新的数列

$$\hat{\gamma}_k = \frac{1}{N}\sum_{j=1}^{N-k} x_j x_{j+k}, \quad k = 0, 1, 2 \cdots \tag{1.2.1}$$

并依次画在图 1.2.1 中,就可以看出它们之间有较明显的差别了. 这又是为什么呢? 这些新的数列又能揭示出原来的随机序列的哪些属性呢? 在本书前五章,我们将逐步回答这些问题. 作为这一切的准备,我们先要介绍描述随机序列的一些必要的概念和简单性质.

一、随机序列的定义

随机序列就是一串随机变量 $\cdots, x_1, x_2, x_3, \cdots$ 所构成的序列,或用 $x_t(t = \cdots, 1, 2, 3, \cdots)$ 表示,对每个固定的整数 t, x_t 是一随机变量.

粗知概率统计概念的读者便知道,若 x 是表示某一随机试验结果的随机变量,那么,做一次试验(或称取一次样),就能获得一个 x 的取值,称为 x 的一个样本值 X,它是一个普通的数,以一定的概率出现. 对于随机序列而言,它的样本值是一串数列 $\cdots X_1$, X_2, X_3, $\cdots$, 它们分别是随机变量 $\cdots x_1$, x_2, x_3, $\cdots$ 各自的样本值.

有三点须要加以说明. 首先应当指出,一个随机变量 x 与它的一个样本值 X,是两种不同意义的量. 前者是对这一具体的随机现象(或随机试验结果)的总称,后者仅是对这个随机现象取得的一个观测值(或试验值). 因此,在上面叙述这两个概念时,前者用小斜体字母 x 和 x_t,后者用大写字母 X 和 X_t 以示区别. 熟悉了这些概念后,自然会分清同一字母 x 或 x_t 何时代表随机变量,何时只是相应的样本值. 所以,本书将随机变量或随机序列与它们

的相应的样本值都用相同的字母表示，这里特加声明.

其次，随机序列 x_t 的整数变量 t，通常代表等间隔的时刻增长量，如用 t 表示第 t 时、第 t 天、第 t 次等. 但是，也可以代表其它的非时间量，比如表示地点、动植物第 t 代等量. 后者非常少见，所以人们常称随机序列为时间序列，而整数变量 t 即认为是指某时刻.

最后还须指出，一个描述随机现象的随机变量，往往可以通过重复试验，获得它的很多个样本值. 用统计方法解决实际问题时，就是要用这些样本值找出被考察的随机现象的各种统计量的值. 但是，当把随机序列概念用于实际现象时，特别是用于 §1 中例 2、例 3 这样的真正随时间而变化的过程时，人们容易想到，这时只可能获得被观察的随机现象的有限数据，即相应的随机序列的一个样本值序列的一段数值，不可能得到一个无穷长度的完整的样本值序列. 当然，得到两个或多个样本值序列更是难以想象的. 这是通常的数理统计与时间序列分析的一个主要不同之点. 今后可以看到，上述不足之处并不妨碍我们利用一段有限样本值来寻找某类随机序列的统计量.

二、随机序列的概率分布

一个随机变量的统计特性完全由它的分布所确定，同样，一个随机序列的统计特性也完全由它的分布所确定. 随机变量（或随机向量）的分布可用分布函数来描述. 随机序列是由无穷多个随机变量构成的，我们说给定了一个随机序列

$$x_t (t = \cdots 1, 2, 3, \cdots)$$

的概率分布，是指对于任意有穷多个时刻 $t_1, t_2, \cdots, t_m$，相应的随机变量 $x_{t_1}, x_{t_2}, \cdots, x_{t_m}$ 的联合分布函数 $F_{t_1 t_2 \cdots t_m}(x_1, x_2, \cdots, x_m)$ 都是被给定的，而且它们之间不能矛盾，即是说由高维联合分布推出的低维联合分布与原给定的低维分布相同.

有时我们也说给定了序列 x_t 的任意有穷维分布.

如果对于任意有穷个不相同的时刻 $t_1, t_2, \cdots, t_m$，相应的

$x_{t_1}, x_{t_2}, \cdots, x_{t_m}$ 是相互独立的随机变量，即

$$F_{t_1t_2\cdots t_m}(x_1, x_2, \cdots, x_m) = F_{t_1}(x_1)F_{t_2}(x_2)\cdots F_{t_m}(x_m) \tag{1.2.2}$$

这时，我们称 x_t 为独立随机序列. 在实际应用中，电路中的热噪声常常近似于这种独立序列.

如果一个随机序列 x_t 的任意有穷维分布具有以下性质：

$$F_{t_1t_2\cdots t_m}(x_1, x_2, \cdots, x_m) = F_{t_1+\tau, t_2+\tau\cdots t_m+\tau}(x_1, x_2, \cdots, x_m) \tag{1.2.3}$$

其中 τ 是任意时刻(τ 是整数)，我们就称 x_t 为狭义平稳序列(或严平稳序列). 也就是说，$x_{t_1}, x_{t_2}, \cdots, x_{t_m}$ 和 $x_{t_1+\tau}, x_{t_2+\tau}, \cdots, x_{t_m+\tau}$ 有相同的分布，无论对怎样的 m 和时刻 $t_1, t_2, \cdots, t_m$ 以及 τ 都如此.

三、随机序列的参数表征

上面已经指出，随机序列的分布决定了随机序列的全部统计特性. 但是，随机序列的分布结构一般是难以完全掌握的，不便于分析、统计和实际应用. 而且，在很多应用中，我们并不需要确定随机序列的分布结构，只要掌握它的某些参数表征就够了. 下面，我们引入随机序列的均值、方差和相关函数等参数表征的定义.

1. 均值函数. 对每个 t 而言，若把随机变量 x_t 的均值记做 $Ex_t = \mu_t$，那么随机序列 x_t 的均值函数就是

$$\mu_t(t = \cdots, 1, 2, 3, \cdots);$$

若 x_t 的分布为 $F_t(x)$，若 x_t 具有密度 $f_t(x)$，则

$$\mu_t = Ex_t = \int x dF_t(x) = \int x f_t(x) dx \tag{1.2.4}$$

有时 μ_t 只取常数，如 §1 中的例 4；有时取周期函数，像 §1 中的例 2；也有时取其它形式. 为了简单，常略去"函数"一词，称 $Ex_t = \mu_t$ 为 x_t 的均值.

2. 自协方差函数. 由(1.2.4)式可知，随机序列的均值只和随机序列的一维分布有关. 为了分析随机序列 x_t 在不同时刻取值的统计关系，须要考虑 x_t 与 x_s 的协方差值. 令

$$\gamma_{ts} = E(x_t - Ex_t)(x_s - Ex_s)$$
$$= \iint (x - \mu_t)(y - \mu_s) f_{ts}(x, y) dx dy \qquad (1.2.5)$$

γ_{ts} 做为 (t, s) 的二元函数，称为随机序列 x_t 的自协方差函数．特别把 $\gamma_{tt} = E(x_t - Ex_t)^2$ 叫做 x_t 的方差函数，或简称方差．

如果一个随机序列 x_t 的任意有穷维分布都是正态分布，我们就称 x_t 为正态随机序列．若以 f 表示相应于 F 的分布密度，此时

$$f_{t_1t_2\cdots t_m}(x_1, x_2, \cdots, x_m) = (2\pi|\Gamma_m|)^{-\frac{m}{2}}$$
$$\cdot \exp\left\{-\frac{1}{2}(\boldsymbol{x}_m - \boldsymbol{\mu}_m)^\tau \Gamma_m^{-1}(\boldsymbol{x}_m - \boldsymbol{\mu}_m)\right\}$$
$$= N(\boldsymbol{\mu}_m, \boldsymbol{\Gamma}_m) \qquad (1.2.6)$$

其中

$$\boldsymbol{x}_m^\tau = (x_1, x_2, \cdots x_m); \qquad \boldsymbol{\mu}_m^\tau = (\mu_{t_1}, \mu_{t_2}, \cdots \mu_{t_m})$$

$$\boldsymbol{\Gamma}_m = \begin{pmatrix} \gamma_{t_1t_1} & \gamma_{t_1t_2} & \cdots & \gamma_{t_1t_m} \\ & \cdots\cdots\cdots & & \\ \gamma_{t_mt_1} & \gamma_{t_mt_2} & \cdots & \gamma_{t_mt_m} \end{pmatrix}$$

很多实际应用的随机序列可近似当做正态序列，对于正态序列，在数学处理上有很多方便之处．

3. 自相关函数．序列的自相关函数 ρ_{ts} 定义为

$$\rho_{ts} = \frac{\gamma_{ts}}{\sqrt{\gamma_{tt}\gamma_{ss}}} \qquad (1.2.7)$$

它刻划了序列 x_t 在不同时刻取值的线性相关程度．

随机序列的参数表征还有很多，与本书关系密切的就是以上三种量，这里就只介绍这些．从 (1.2.4)，(1.2.5)，(1.2.7) 三式易见，μ_t，γ_{ts} 和 ρ_{ts} 被 x_t 的分布唯一决定．但是，反之由 μ_t，γ_{ts} 和 ρ_{ts} 一般并不能唯一确定 x_t 的分布．也就是说，具有不同分布的随机序列可以有相同的均值、自协方差和自相关函数．虽然如此，对于大量的实际应用而言，通过以上三种量来掌握随机序列的统计特性已经够用了．当然，在处理某些实际问题时，也不排除须要了

解随机序列的分布情况，不过，这不是我们讨论的内容。

§3. 平稳随机序列

为了便于读者掌握，我们把本书的讨论几乎完全限于正态序列范围之内，这不会影响时序分析方法的介绍，而且会使很多数学概念和性质有比较简单的形式. 只在某些个别情况下，我们指出对于非正态序列的类似结果. 此外，本书所介绍的各种方法是以广义平稳序列为基础的，象前面所定义的狭义平稳序列的一般性质，我们不去深入研究. 下面叙述广义平稳序列的定义和性质.

一、广义平稳序列的定义

若随机序列 x_t 的二阶矩有穷而且对任意时刻 t 和 s 满足

$$Ex_t = \mu \tag{1.3.1}$$

$$Ex_tx_s - \mu^2 = \gamma_{ts} = \gamma_{t-s} \tag{1.3.2}$$

则称它为广义平稳序列(或宽平稳序列)，即 μ_t 与 t 无关，γ_{ts} 只与 $s-t$ 有关. “广义”是相对(1.2.3)式的“狭义”而言的，如果狭义平稳序列 x_t 的一、二阶矩都有穷，它一定也是广义平稳的，反之则不然. 如果 x_t 是正态随机序列，那么从(1.2.6)式易知，x_t 的狭义平稳性和广义平稳性是一致的. 在很多书籍中，都把广义平稳过程简称为平稳过程，我们也用“平稳序列”的简称来代替“广义平稳序列”一词. 在实际应用时，平稳序列仅仅是对于真实随机序列的一种近似描述手段，比如电路中的热噪声，陀螺仪的漂移速率，以及其它精密仪表的漂移误差等，常用平稳序列(或平稳过程)描述. 至于一个随机序列在什么情况下才可以用平稳序列近似描述，本书第三章将给出一种粗略的辨别方法.

二、自协方差函数与谱分布

依前所述，对于正态平稳序列 x_t，其均值 μ 和自协方差函数 γ_{s-t} 唯一决定了它的分布，从而也就决定了它的全部统计性质. 可见，自协方差函数 γ_{s-t} 起着非常重要的作用. 关于 γ_{s-t} 的性质

有以下两点:

1. 对称性: 即 $\gamma_\tau = \gamma_{-\tau}$, 这是因为

$$\gamma_\tau = E x_{t-\tau} x_t = E x_t x_{t-\tau} = \gamma_{-\tau} \tag{1.3.3}$$

2. 非负定性: 即对任意正整数 m, 方阵

$$\Gamma_m = \begin{pmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{m-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{m-2} \\ \cdots & \cdots & \cdots & \cdots \\ \gamma_{m-1} & \gamma_{m-2} & \cdots & \gamma_0 \end{pmatrix} \tag{1.3.4}$$

是非负定的,也就是说,对于任意 m 维实值非零向量 $\xi = (\xi_0, \xi_1, \cdots, \xi_{m-1})^\tau$, 都有

$$\xi^\tau \Gamma_m \xi = \sum_{i,j=0}^{m-1} \gamma_{i-j} \xi_i \xi_j \geqslant 0 \tag{1.3.5}$$

这是因为

$$\sum_{i,j=0}^{m-1} \gamma_{i-j} \xi_i \xi_j = \sum_{i,j=0}^{m-1} \xi_i \xi_j E(x_i - \mu)(x_j - \mu)$$

$$= E\left[\sum_{i=0}^{m-1} \xi_i (x_i - \mu)\right]^2 \geqslant 0 \tag{1.3.6}$$

有时,我们把满足上述性质 1 和 2 的实数列 $\gamma_0, \gamma_1, \gamma_2, \cdots$ 称为非负定列. 根据(1.3.5)式不难看出, $\gamma_0 \geqslant 0$ 及 $|\gamma_\tau| \leqslant \gamma_0$. 反之,任意一个非负定列必为某平稳序列的自协方差函数(见[3]),而且有以下定理.

定理 1. 设 γ_k 为一平稳序列的自协方差函数,则存在一有界非降函数 $G(\lambda)$, 使得

$$\gamma_k = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{i2\pi k\lambda} dG(\lambda) \tag{1.3.7}$$

$G(\lambda)$ 称为平稳序列 x_t 的谱分布. 如果 $G(\lambda)$ 可以微分,并记

$$g(\lambda) = \frac{dG(\lambda)}{d\lambda}$$

那么

$$\gamma_k = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{i2\pi k\lambda} g(\lambda) d\lambda \tag{1.3.8}$$

$g(\lambda)$ 称为平稳序列 x_t 的谱密度. 当 $\sum_{k=-\infty}^{\infty}|\gamma_k| < \infty$ 时, $g(\lambda)$ 一定存在,而且

$$g(\lambda) = \sum_{k=-\infty}^{\infty} \gamma_k e^{-i2\pi k\lambda} \tag{1.3.9}$$

在一些工程方面的书上, $g(\lambda)$ 也被称为序列的功率谱密度.

三、白噪声序列

如果平稳序列 a_t 的均值为零,自协方差函数 $\gamma_k = \sigma_a^2 \delta_{k0}$, 我们称这样的 a_t 为白噪声序列,或简称白噪声. 它的谱密度

$$g_a(\lambda) = \sum_{k=-\infty}^{\infty} \sigma_a^2 \delta_{k0} e^{-i2\pi k\lambda} = \sigma_a^2 \tag{1.3.10}$$

可见 $g_a(\lambda)$ 为一常数, 即序列 a_t 的谱密度在各个频率上具有相同的分量. 正象白光一样, 等量地包含了各种有色光的光频分量. 在自然界里,纯粹的白噪声很难遇到,但是很多实际的随机序列可以近似地符合白噪声的性质. 白噪声在时序分析理论与方法的研究中起着十分重要的作用.

四、平稳序列的线性运算

随机变量可以进行加减等运算, 随机序列也如此. 比如 x_t 是一平稳序列, α, β 是两个实数, τ 是某一固定时刻, 那么

$$y_t = \alpha x_t + \beta x_{t-\tau} \tag{1.3.11}$$

仍然是平稳序列. 而且这一运算可以推广到无穷求和的情况, 即假定 α_k 是实数列, 且 $\sum_{k=-\infty}^{\infty}|\alpha_k| < \infty$, 那么容易验证

$$y_t = \sum_{k=-\infty}^{\infty} \alpha_k x_{t-k} \tag{1.3.12}$$

也是平稳序列, 其中 $\sum_{k=-\infty}^{\infty} \alpha_k x_{t-k}$ 作为当 $M, N \to \infty$ 时 $\sum_{k=-M}^{N} \alpha_k x_{t-k}$ 的均方极限.

注. 设 $\{x_k\}$ 为平稳序列. 若 $E|x_k - x|^2 \xrightarrow[k\to\infty]{} 0$, 称 $\{x_k\}$ 均方收敛于 x. 又称 x 为 $\{x_k\}$ 的均方极限.

一类非常有用的特殊情况是取 $x_t = a_t$ 为白噪声, $\alpha_k = 0$, 当 $k < 0$, 这时, 平稳序列

$$y_t = \sum_{k=0}^{\infty} \alpha_k a_{t-k} \tag{1.3.13}$$

称为 a_t 的滑动和;特别若再有 $\alpha_k = 0$, 当 $k > q$, 则

$$y_t = \sum_{k=0}^{q} \alpha_k a_{t-k} \tag{1.3.14}$$

称为 a_t 的 q 阶滑动平均.

由以上两式所给出的平稳序列 y_t, 其自协方差函数和谱密度可以利用 a_t 的性质很方便地求出. 以(1.3.13)的 y_t 为例,约定当 $k < 0$ 时, $\alpha_k = 0$, 则

$$\begin{aligned}\gamma^y_{s-t} &= E\left(\sum_{k=0}^{\infty} \alpha_k a_{t-k} \sum_{j=0}^{\infty} \alpha_j a_{s-j}\right) = \sum_{k,j=0}^{\infty} \alpha_k \alpha_j E a_{t-k} a_{s-j} \\ &= \sigma_a^2 \sum_{k,j=0}^{\infty} \alpha_k \alpha_j \delta_{s-t+k-j,0} = \sigma_a^2 \sum_{k=0}^{\infty} \alpha_k \alpha_{k+s-t} \\ &= \sigma_a^2 \sum_{k=|s-t|}^{\infty} \alpha_k \alpha_{k-|s-t|}\end{aligned} \tag{1.3.15}$$

$$\begin{aligned}g_y(\lambda) &= \sum_{\tau=-\infty}^{\infty} \gamma^y_\tau e^{-i2\pi\tau\lambda} = \sigma_a^2 \sum_{\tau=-\infty}^{\infty} \sum_{k=0}^{\infty} \alpha_k \alpha_{k+\tau} e^{-i2\pi\tau\lambda} \\ &= \sigma_a^2 \sum_{k=0}^{\infty} \sum_{\tau=-k}^{\infty} \alpha_k \alpha_{k+\tau} e^{-i2\pi\tau\lambda} = \sigma_a^2 \sum_{k=0}^{\infty} \sum_{\tau=-k}^{\infty} \alpha_k e^{i2\pi k\lambda} \alpha_{k+\tau} \\ &\quad \cdot e^{-i2\pi(k+\tau)\lambda} = \sigma_a^2 \left| \sum_{k=0}^{\infty} \alpha_k e^{-i2\pi k\lambda} \right|^2\end{aligned} \tag{1.3.16}$$

若令

$$A(\omega) = \sum_{k=0}^{\infty} \alpha_k \omega^k$$

则(1.3.16)可表为

$$g_y(\lambda) = \sigma_a^2 \left| A(e^{-i2\pi\lambda}) \right|^2 \tag{1.3.17}$$

由此可见，虽然白噪声 a_t 具有最简单的常值谱密度，但是，由它的线性运算所得到的 y_t 可以具有比较复杂的连续谱密度，其中包括下面将要介绍的有理谱密度。

五、具有有理谱的平稳序列

如上所述，对于正态平稳序列，只要知道了它的自协方差函数 γ_k，或者知道了它的谱分布 $G(\lambda)$，就等于掌握了它的统计性质。主要利用前者去分析时间序列时，称为"相关分析法"或"时域分析法"；主要利用后者时，称为"频谱分析法"。怎样求得一个正态平稳序列 x_t 的 γ_k 或 $G(\lambda)$ 呢？除了从物理上进行某些分析之外，主要是根据 x_t 的样本值 $x_1, x_2, \cdots, x_N$ 对 γ_k 或 $G(\lambda)$ 进行估计，这恰是时序分析要解决的主要问题之一。但是，这里存在两个难点：首先，自协方差函数 $\gamma_k(k = 0, 1, 2, \cdots)$ 是由无穷多个值构成的，谱函数 $G(\lambda)$ 则为在 $\left[-\frac{1}{2}, \frac{1}{2}\right]$ 内取值的函数，要用有穷个 x_t 的样本值对所有 γ_k 或 $G(\lambda)$ 的所有取值进行估计，这就是一个难点。其次，即使能够对 γ_k 或 $G(\lambda)$ 的所有取值做出估计，由于 $\gamma_0, \gamma_1, \cdots$ 或函数 $G(\lambda)$ 的形状复杂，也不利于在预报、控制或模拟等应用中使用。为了克服这两个困难，我们采用绪论中提到的"参数化"的方法，即把 $G(\lambda)$ 局限在一个较窄的函数范围内进行讨论。具体地说，我们只讨论这样一类正态平稳序列，它们的谱分布 $G(\lambda)$ 不仅可微，而且它的导函数（即谱密度）$g(\lambda) = \frac{dG(\lambda)}{d\lambda}$ 为 $e^{-i2\pi\lambda}$ 的有理函数：

$$g(\lambda) = \sigma_a^2 \left| \frac{\theta(e^{-i2\pi\lambda})}{\varphi(e^{-i4\pi\lambda})} \right|^2 \tag{1.3.18}$$

其中 $\theta(\omega)$ 和 $\varphi(\omega)$ 为 ω 的实系数多项式：

$$\theta(\omega) = 1 - \theta_1\omega - \theta_2\omega^2 - \cdots - \theta_q\omega^q \tag{1.3.19}$$

$$\varphi(\omega) = 1 - \varphi_1\omega - \varphi_2\omega^2 - \cdots - \varphi_p\omega^p \tag{1.3.20}$$

两者无公共因子，并且限定 $\theta(\omega)$ 和 $\varphi(\omega)$ 的根全在复平面的单位圆外. 这样一来，为了估计 $G(\lambda)$，只要估计 p, q, σ_a^2 和 (1.3.19), (1.3.20) 式的诸系数就可以了. 而这些只不过是有限个参数而已. 有了这些参数的估计值，利用 (1.3.18) 式就可得出 $g(\lambda)$ 在 $\left[-\frac{1}{2}, \frac{1}{2}\right]$ 上的各处取值的估计. 在应用于预报、控制和模拟等目的时，由于认为 $g(\lambda)$ 具有 (1.3.18) 式的形式，解决问题时就容易下手了.

谱密度有 (1.3.18) 这种形式的 x_t，称为具有有理谱的平稳序列. 还须指出，自然现象中的某些随机序列虽然不一定真的具有有理谱，但是，反映实际物理过程的平稳序列，绝大多数都存在谱密度，而且在很多情况下还可以认为它是连续的. 因此，在大量的实际应用中，用有理谱来逼近真实的谱是一种合理的方法.

六、随机差分方程

根据前面的叙述，具有有理谱的平稳序列的自协方差函数也被以上诸参数所决定. 事实上，利用定理 1 便知

$$\gamma_k = \sigma_a^2\int_{-\frac{1}{2}}^{\frac{1}{2}}\left|\frac{\theta(e^{-i2\pi\lambda})}{\varphi(e^{-i2\pi\lambda})}\right|^2 e^{i2\pi k\lambda}d\lambda,\quad k=0,1,2,\cdots \tag{1.3.21}$$

反之，若 γ_k 能表示成 (1.3.21) 式的形式，则随机序列也一定具有 (1.3.18) 形式的有理谱密度.

现在举一个简单例子，取

$$g(\lambda)=\frac{\sigma_a^2}{|1-\varphi e^{-i2\pi\lambda}|^2},\quad |\varphi|<1 \tag{1.3.22}$$

则

$$\begin{aligned}\gamma_k &= \sigma_a^2\int_{-\frac{1}{2}}^{\frac{1}{2}}\frac{1}{|1-\varphi e^{-i2\pi\lambda}|^2}\,e^{i2\pi k\lambda}d\lambda\\ &= \sigma_a^2\int_{-\frac{1}{2}}^{\frac{1}{2}}\frac{\cos 2\pi k\lambda}{1-2\varphi\cos 2\pi\lambda+\varphi^2}\,d\lambda\end{aligned}$$

由此算出：

$$\gamma_k = \frac{\sigma_a^2}{1-\varphi^2}\varphi^k,\ k=0,1,2\cdots \tag{1.3.23}$$

这表明自协方差函数为一指数型数列．反之，若 γ_k 具有（1.3.23）式的形式，则相应的随机序列一定具有（1.3.22）式的谱密度．

虽然（1.3.21）式反映了有理谱与其相应的自协方差函数之间的关系，但是，为了以后的时域分析，还要引进随机差分方程的概念．设 $\varphi_1,\varphi_2,\cdots,\varphi_p$ 和 $\theta_1,\theta_2,\cdots,\theta_q$ 分别是 p 个和 q 个实数，并设以它们为系数的两个多项式 $\varphi(\omega)$ 和 $\theta(\omega)$（见(1.3.19)和(1.3.20)式）无公共因子，且它们的根全在单位圆外．如果平稳序列 x_t 满足如下关系式：

$$x_t-\varphi_1x_{t-1}-\cdots-\varphi_px_{t-p}=a_t-\theta_1a_{t-1}-\cdots-\theta_qa_{t-q},\quad t=\cdots-1,0,1\cdots \tag{1.3.24}$$

其中 a_t 是一白噪声，$Ea_t^2=\sigma_a^2$，且当 $s>t$ 时，$Ex_ta_s=0$，那么，我们就说 x_t 是随机差分方程(1.3.24)的一个平稳解．根据平稳序列的理论(见附录，§1)可知，具有有理谱密度(1.3.18)式的平稳序列，一定是随机差分方程(1.3.24)式的一个平稳解；反之，(1.3.24)式的平稳解一定具有(1.3.18)式的谱密度．这样，就在具有有理谱的平稳序列与随机差分方程的平稳解之间建立了一一对应的关系．这种关系比(1.3.21)式更直截了当，也更便于运算和应用．仍以具有谱密度(1.3.22)式的平稳序列为例，由这里建立的关系，它应满足下面的差分方程：

$$x_t-\varphi x_{t-1}=a_t \tag{1.3.25}$$

于是有

$$E(x_{t-k}x_t-\varphi x_{t-k}x_{t-1})=Ex_{t-k}a_t=0,\ k>0$$

即

$$\gamma_k=\varphi\gamma_{k-1},\qquad k>0$$

由(1.3.25)式又有

$$\gamma_0=Ex_t^2=E(a_t+\varphi x_{t-1})^2=\sigma_a^2+\varphi^2\gamma_0$$

所以 $\gamma_k=\varphi\gamma_{k-1}=\varphi^2\gamma_{k-2}=\cdots=\varphi^k\gamma_0=\dfrac{\sigma_a^2}{1-\phi^2}\varphi^k$，这与

(1.3.23)式完全一样，但计算方便．不仅如此，(1.3.25)式还清楚地表明了 x_t 的前后依赖关系．事实上，(1.3.25)式又可写为

$$x_t = \varphi x_{t-1} + a_t$$

熟悉回归分析的读者立刻看出，这是一种形如

$$x_t = \varphi y_t + a_t$$

的回归方程，不过，这时 $y_t = x_{t-1}$ 是序列 x_t 自身的滞后一步．所以(1.3.25)式的平稳解 x_t 又称为一阶自回归序列，而参数 φ 则表示 x_t 前后的相关程度．此外，由(1.3.25)式还可推知

$$\begin{aligned}x_t &= \varphi x_{t-1} + a_t = \varphi(\varphi x_{t-2} + a_{t-1}) + a_t \\ &= \varphi^2 x_{t-2} + \varphi a_{t-1} + a_t = \cdots = \varphi^n x_{t-n} \\ &\quad + \varphi^{n-1} a_{t-n+1} + \varphi^{n-2} a_{t-n+2} + \cdots + a_t\end{aligned}$$

由于 $\varphi(\omega) = 1 - \varphi\omega$ 的根在单位圆外(因 $|\varphi| < 1$)，所以从上式取极限得到

$$x_t = \lim_{n\to\infty}\left(\varphi^n x_{t-n} + \sum_{j=0}^{n-1}\varphi^j a_{t-j}\right) = \sum_{j=0}^{\infty}\varphi^j a_{t-j}$$

这恰是形如(1.3.13)式的滑动和．注意，x_t 只是 $a_t, a_{t-1}, \cdots$ 的滑动和，而 a_t 是白噪声，所以当 $t<s$ 时，x_t 与 a_s 独立，即 $Ex_ta_s = 0$．下面再从滑动和回到谱密度，因 $|\varphi| < 1$，由(1.3.17)式知

$$A(\omega) = \sum_{j=0}^{\infty}\varphi^j\omega^j = \frac{1}{1-\varphi\omega}, \qquad |\omega| \leqslant 1$$

$$g(\lambda) = \sigma_a^2 |A(e^{-i2\pi\lambda})|^2 = \frac{\sigma_a^2}{|1-\varphi e^{-i2\pi\lambda}|^2}$$

$$= \frac{\sigma_a^2}{1+\varphi^2 - 2\varphi\cos 2\pi\lambda}$$

从此以后，我们所讨论的平稳随机序列，不仅限于正态序列，而且都具有有理谱密度．换言之，它们必是(1.3.24)型的随机差分方程的平稳解．利用(1.3.24)式，我们将在下一章里详细分析相应的 γ_k 的各种性质．

七、遍历性

在给出具有有理谱的平稳序列的定义以后，我们要特别指出，实际中所取随机序列 x_t 的样本数据总是有限个，要用这有限长度的样本序列值，来估计上面所说的诸参数 p，q，σ_a^2 和 $\varphi(\omega)$，$\theta(\omega)$ 的系数等值，常用的统计方法中的估计手段就不够用了．而应用新手段的一个重要前提就是随机过程要具有遍历性．

遍历性定义． 设 x_t 为一随机序列，$v(x_t)$ 是 x_t 的函数（如 $|x_t|$，x_t^2，$x_tx_{t+\tau}$ 等等），如果对于任何使 $Ev(x_t)$ 存在的函数 v，概率为一地有

$$Ev(x_t)=\lim_{N\to\infty}\frac{1}{N}\sum_{j=1}^{N}v(x_{t+j}) \qquad (1.3.26)$$

则称 x_t 为具有遍历性的随机序列．

对于本书主要讨论的正态有理谱平稳序列，根据随机过程理论(见[5])可知，它们一定具有遍历性．至于有关遍历性定义的其它一些形式，我们这里不予详论．

遍历性的物理意义可以大致叙述如下：随机序列 x_t 的函数 $v(x_t)$ 也是一个随机变量，其均值为 $Ev(x_t)$，这可以称之为 $v(x_t)$ 的总体平均（即依 $v(x_t)$ 的分布所求出的均值，也有人叫它相平均）．又当 t 固定，而将 $v(x_{t+j})$，$j=1,2,3,\cdots$ 视为一个随机序列时，$\lim\limits_{N\to\infty}\frac{1}{N}\sum\limits_{j=1}^{N}v(x_{t+j})$ 称为 $v(x_t)$ 的时域平均．所谓 x_t 的遍历性，简而言之，就是对任何函数 v，$v(x_t)$ 的总体平均等于它的时域平均，粗略地说，这意味着 x_{t+j} 的任何一个样本随着 j 的变化所能取的值，依随机变量 x_t 的概率分布，历经它所能取的各种值．这里顺便指出，若 x_t 具有遍历性，它的线性运算也具有这一性质．

最后仅举一例说明遍历性的用处．取 $v(x_t)=x_tx_{t+1}$，则由遍历性，

$$\gamma_1=Ex_tx_{t+1}=\lim_{N\to\infty}\frac{1}{N}\sum_{j=1}^{N}x_{t+j}x_{t+1+j}$$

这说明，当N很大时，上式右边的平均值可以做为γ_1的近似估计值.

§4. 多维随机序列

本章一开始所举的四个例子，不论是随机过程还是非随机过程，它们所牵涉的量都是标量的过程，即一维变量. 在某些情况下，我们须要同时考察几种过程量，要分析它们彼此之间和前后之间的相互联系. 例如，在进行雨量预报研究时，就须要同时分析该地区的雨量、气温和气压等. 又如在炼钢过程中，钢水的温度和炉气的二氧化碳含量都与钢的脱碳量有关，为了更好地在线定碳，也应当考虑多种有关的量.

一维随机序列的定义不难推广到多维情形. 我们称$\boldsymbol{y}_t$为一r维随机序列，是指对每个固定的整数值时刻t而言，$\boldsymbol{y}_t$是一r维随机向量，常常记做$\boldsymbol{y}_t^\tau=(y_t^{(1)}, y_t^{(2)}, \cdots, y_t^{(r)})$，时刻$t$可以从$-\infty$变到$+\infty$. 这$r$个随机序列$y_t^{(1)}, \cdots, y_t^{(r)}$相互之间有一定的统计联系，为了描述这些联系，还要以下一些推广的概念.

一、多维随机序列的均值函数

对于多维序列$\boldsymbol{y}_t$来说，固定t时，$E\boldsymbol{y}_t$不再是一个实数，而是一个向量，即

$$E\boldsymbol{y}_t=(Ey_t^{(1)}, Ey_t^{(2)}, \cdots, Ey_t^{(r)})^\tau \tag{1.4.1}$$

当t变化时，$E\boldsymbol{y}_t$是一个（非随机的）r维向量序列，它称为$\boldsymbol{y}_t$的均值函数，有时也简称均值.

二、方差阵、自协方差阵与互协方差函数

对于随机向量序列$\boldsymbol{y}_t$的每个分量$y_t^{(k)}(k=1, 2, \cdots, r)$，$E(y_t^{(k)}-Ey_t^{(k)})^2$是$y_t^{(k)}$的方差函数，但是，用这$r$个方差函数还不能完全反映出$\boldsymbol{y}_t$"方差"的全貌. 因此我们定义

$$E(\boldsymbol{y}_t-E\boldsymbol{y}_t)(\boldsymbol{y}_t-E\boldsymbol{y}_t)^\tau \tag{1.4.2}$$

为$\boldsymbol{y}_t$的方差阵函数；这是一个r阶非负定矩阵，它的主对角线上

的元恰是 $y_t^{(k)}$ 的方差，而 i 行 j 列的元则为 $y_t^{(i)}$ 与 $y_t^{(j)}$ 的互协方差.

为了掌握 $\boldsymbol{y}_t$ 在不同时刻取值的统计关系，我们还定义

$$R_{ts} = E(\boldsymbol{y}_t - E\boldsymbol{y}_t)(\boldsymbol{y}_s - E\boldsymbol{y}_s)^\tau \tag{1.4.3}$$

为 $\boldsymbol{y}_t$ 的自协方差阵函数；它的主对角线元素是 $y_t^{(k)}(k=1,2,\cdots,r)$ 的自协方差函数，而 i 行 j 列的元 $E(y_t^{(i)} - Ey_t^{(i)})(y_s^{(j)} - Ey_s^{(j)})$ 称为 $y_t^{(i)}$ 与 $y_t^{(j)}$ 的互协方差函数. 当 $t=s$ 时，R_{tt} 就是 $\boldsymbol{y}_t$ 的方差阵函数. 由此可见，R_{ts} 更进一步揭示了 $\boldsymbol{y}_t$ 的各分量间及前后之间的相互联系.

三、多维平稳序列

设 $\boldsymbol{y}_t$ 为 r 维随机序列，如果它还满足

$$E\boldsymbol{y}_t = \boldsymbol{\mu};\quad R_{ts} = E(\boldsymbol{y}_t - \boldsymbol{\mu})(\boldsymbol{y}_s - \boldsymbol{\mu})^\tau = R_{t-s} \tag{1.4.4}$$

那么就称 $\boldsymbol{y}_t$ 为 r 维平稳随机序列.

有关一维平稳序列的概念和性质，同样可以推广到多维平稳序列，我们不再一一列举. 这里，只简单介绍一下具有有理谱的多维平稳序列，因为它比一维情形稍微复杂一点. 同一维情形类似，在具有有理谱的多维平稳序列与某类多维随机差分方程的平稳解之间，存在着相互对应的关系. 我们将只引进后者. 设 $A_1,\cdots,A_p$ 和 $B_1,\cdots,B_q$ 都是 r 阶实方阵，定义以它们为系数的两个矩阵多项式：

$$A(\omega) = I - \omega A_1 - \omega^2 A_2 - \cdots - \omega^p A_p \tag{1.4.5}$$

$$B(\omega) = I - \omega B_1 - \omega^2 B_2 - \cdots - \omega^q B_q \tag{1.4.6}$$

并设 $A(\omega)$ 与 $B(\omega)$ 为左既约的，即若 $A(\omega) = P(\omega)A_1(\omega)$，$B(\omega) = P(\omega)B_1(\omega)$，则 $|P(\omega)|$ 为常数，而且 $\operatorname{rank}(A_p, B_q) = r$，此外，$|A(\omega)| = 0$，$|B(\omega)| = 0$ 的根全在单位圆外. 如果 r 维平稳序列 $\boldsymbol{y}_t$ 满足如下方程：

$$\begin{aligned}\boldsymbol{y}_t &- A_1\boldsymbol{y}_{t-1} - \cdots - A_p\boldsymbol{y}_{t-p}\\ &= \boldsymbol{v}_t - B_1\boldsymbol{v}_{t-1} - \cdots - B_q\boldsymbol{v}_{t-q}\end{aligned} \tag{1.4.7}$$

其中 $\boldsymbol{v}_t$ 是 r 维白噪声，即满足 $E\boldsymbol{v}_t = 0$，$E\boldsymbol{v}_t\boldsymbol{v}_s^\tau = Q_v\delta_{ts}$，又 $t < s$

时，$Ey_t v_t^{\tau}=0$，那么就说 y_t 是 r 维随机差分方程（1.4.7）的一个平稳解.

§5. 两种估计及参数估计的优效性概念

一、最小二乘法

最小二乘法是大家所熟知的一种估计方法. 特别，线性参数的最小二乘法是常用的估计方法之一. 这里主要介绍非线性参数的最小二乘法，这种方法在第四章中将要用到. 我们先扼要叙述一下线性参数的最小二乘法.

假定有一线性模型

$$y_k=a_{k1}\beta_1+a_{k2}\beta_2+\cdots+a_{kr}\beta_r+e_k,\quad k=1,2,\cdots \tag{1.5.1}$$

其中 $a_{k1},a_{k2},\cdots,a_{kr}(k=1,2,\cdots)$ 是已知量，$\beta_1,\beta_2,\cdots,\beta_r$ 是未知参数，$e_k(k=1,2,\cdots)$ 称为模型（1.5.1）的残差，它是在零上下起伏的量，且与 $a_{kj}(j=1,\cdots,r,\ k=1,2,\cdots)$ 无关. 如果通过测量后，获得 y_k 的 n 个数据 $y_1,y_2,\cdots,y_n$，那么，使得残差平方和

$$S(\beta_1,\beta_2,\cdots,\beta_r)=\sum_{k=1}^{n}(y_k-a_{k1}\beta_1-a_{k2}\beta_2-\cdots-a_{kr}\beta_r)^2$$

$$=\sum_{k=1}^{n}e_k^2 \tag{1.5.2}$$

达到极小的解 $\hat{\beta}_1,\hat{\beta}_2,\cdots,\hat{\beta}_r$ 称为参数 $\beta_1,\beta_2,\cdots,\beta_r$ 的最小二乘估计. 很容易证明（见[5]）其解为

$$\hat{\boldsymbol{\beta}}=(A^{\tau}A)^{-1}A^{\tau}\boldsymbol{y} \tag{1.5.3}$$

其中 $\hat{\boldsymbol{\beta}}^{\tau}=(\hat{\beta}_1,\hat{\beta}_2,\cdots,\hat{\beta}_r)$，$A$ 为 $a_{kj}(j=1,2,\cdots,r,\ k=1,2,\cdots,n)$ 组成的 $n\times r$ 矩阵，$\boldsymbol{y}^{\tau}=(y_1,y_2,\cdots,y_n)$.

现在，我们把上面的模型加以推广. 考虑模型

$$y_k=f_k(\boldsymbol{\beta};y_1,\cdots,y_{k-1})+e_k,\quad k=1,2,\cdots \tag{1.5.4}$$

其中 e_k 仍然表示残差，$\boldsymbol{\beta}^{\tau}=(\beta_1,\beta_2,\cdots,\beta_r)$ 为未知参数矢量，f_k 是 $(\beta,y_1,\cdots,y_{k-1})$ 的函数，对 $\boldsymbol{\beta}$ 非线性. 如果获得了测量值 $y_1,y_2,\cdots,y_n$，那么，使得残差平方和

$$S(\boldsymbol{\beta}) = \sum_{k=1}^{n} [y_k - f_k(\boldsymbol{\beta}; y_1, \cdots, y_{k-1})]^2 \qquad (1.5.5)$$

达到极小的解 $\hat{\boldsymbol{\beta}}$，即称为 $\boldsymbol{\beta}$ 的最小二乘估计，以后简称 LS 估计(Least Square)．一般说来，对非线性参数而言，$\hat{\boldsymbol{\beta}}$ 的求解比线性情形要麻烦得多，且只能给出数值解法，而得不到如(1.5.3)式这样的明显解．此外，f_k 还可能是用迭代方式给出，而不必有明显的函数表达式．有了这些推广以后，很多复杂的参数估计问题都可以化成最小二乘估计问题．这一点将在第四章§2中看到．

在这里，我们有意不涉及最小二乘估计的统计特性，这是为了使不熟悉概率统计知识的读者也能掌握和运用这一方法．但是，为要分析估计 $\hat{\boldsymbol{\beta}}$ 的误差情况，那就应当引入 e_k 和 y_k 的统计特性，通常都假定 e_k 为白噪声，而且 e_k 与 $y_1, \cdots, y_{k-1}$ 独立．这时，做为 $y_1, \cdots, y_n$ 的函数的最小二乘估计 $\hat{\boldsymbol{\beta}}$ 和真值 $\boldsymbol{\beta}$ 之间的接近程度，可用本章后面所介绍的几种估计量优效性来衡量，在本书附录§5中，具体分析某些最小二乘估计量的各种优效性质．

二、最小方差估计

最小方差估计，又简称LMS估计(Least Mean Square)，我们先对一类特殊情形的最小方差估计做一简单介绍，并且讨论它的一些性质，因为在本书第四、七、九等章经常用到它们．

设 $w_k(M-1<k<N+1, N, M$ 为整数或正负无穷)是一组正态随机变量，并且它们的均值都为零．又设正态随机变量 z 的均值也为零，而且 z 与 w_k 的联合分布也是正态分布．所谓根据 $w_k(M-1<k<N+1)$ 对 z 做的(或 z 关于 w_k 的)最小方差估计 $\hat{z}$，是指存在如下的量：

$$\hat{z} = \sum_{k=M}^{N} \beta_k w_k \qquad (1.5.6)$$

其中系数 β_k 使误差方差 $E(z-\hat{z})^2$ 达到极小，即

$$E(z-\hat{z})^2 = \inf_{a_k} E\left(z - \sum_{k=M}^{N} a_k w_k\right)^2 \qquad (1.5.7)$$

我们把这种估计记做 $\hat{z} = E(z|w_k, M-1<k<N+1)$.

对于我们来说，最有用的情况是：每一 w_k 可以表成白噪声 $a_j(M-1<j\leqslant k)$ 的和，即

$$w_k = \sum_{j=M}^{k} \xi_{kj} a_j \tag{1.5.8}$$

同时每一 a_j 也能表成 $w_l(M-1<l\leqslant j)$ 的和，即

$$a_j = \sum_{l=M}^{j} \eta_{jl} w_l \tag{1.5.9}$$

且其中的系数 ξ_{kj}, η_{jl} 又满足如下条件：

$$|\xi_{kj}| \leqslant g_1 e^{-g_2(k-j)}, \qquad |\eta_{jl}| \leqslant g_1 e^{-g_2(j-l)} \tag{1.5.10}$$

此处 g_1 和 g_2 表示与 k, j, l 无关的正实数. 当 M, N 为有穷整数时，上述这些条件是自然满足的；而当 $M=-\infty$ 时，在第二章和附录中将会看到，如果 w_k 是(1.3.24)式的平稳解，那末上述条件也能满足. 对于这样的 w_k，随机变量 z 的最小方差估计有较简便的形式，因而较易于讨论它们的性质.

我们先引进两个随机变量集合：

$$\mathscr{A} = \left\{y: y = \sum_{j=M}^{N} \beta_j a_j,\ \beta_j \text{ 为实数},\ \sum_{j=M}^{N} \beta_j^2 < \infty\right\} \tag{1.5.11}$$

$$\mathscr{W} = \left\{y: y = \sum_{l=M}^{N} \alpha_l w_l,\ \alpha_l \text{ 为实数},\ \sum_{l=M}^{N} \alpha_l^2 < \infty\right\} \tag{1.5.12}$$

然后从它们出发来讨论最小方差估计的性质.

1. $\mathscr{A} = \mathscr{W}$. 当 M, N 都为有穷时，这是显然的，我们只对 $M=-\infty$, $N=+\infty$ 的情况加以证明. 若 $y \in \mathscr{A}$，则存在一串 β_j，使得 $\sum\limits_{j=-\infty}^{\infty} \beta_j^2 < \infty$，且

$$y = \sum_{j=-\infty}^{\infty} \beta_j a_j$$

由(1.5.9)式，

$$y = \sum_{j=-\infty}^{\infty} \beta_j \sum_{l=-\infty}^{j} \eta_{jl} w_l = \sum_{l=-\infty}^{\infty} \left(\sum_{j=l}^{\infty} \beta_j \eta_{jl}\right) w_l = \sum_{l=-\infty}^{\infty} \alpha_l w_l$$

其中 $\alpha_l = \sum_{j=l}^{\infty} \beta_j \eta_{jl}$. 又由(1.5.10)式和 Schwarz 不等式可知

$$\sum_{l=-\infty}^{\infty} \alpha_l^2 = \sum_{l=-\infty}^{\infty} \sum_{j,k=l}^{\infty} \beta_j \beta_k \eta_{jl} \eta_{kl} \leqslant \sum_{l=-\infty}^{\infty} \sum_{j,k=l}^{\infty} |\beta_j \beta_k| g_1^2 e^{-g_2(j-l+k-l)}$$

$$= \sum_{m,n=0}^{\infty} \left(\sum_{l=-\infty}^{\infty} |\beta_{m+l} \beta_{n+l}| \right) g_1^2 e^{-g_2(m+n)}$$

$$\leqslant \left(\sum_{l=-\infty}^{\infty} \beta_l^2 \right) \left(\frac{g_1}{1 - e^{-g_2}} \right)^2 < \infty$$

因此 $y \in \mathscr{W}$. 因为 y 是 $\mathscr{A}$ 的任意元，所以 $\mathscr{W} \supset \mathscr{A}$. 用同样方法可以证明 $\mathscr{A} \supset \mathscr{W}$，因此 $\mathscr{A} = \mathscr{W}$. 熟悉泛函分析的读者容易看出，$\mathscr{A} = \mathscr{W}$ 是 Hilbert 空间，估计量 $\hat{z} = E(z | w_k, M-1 < k < N+1)$ 是 z 在 $\mathscr{W}$ 上的投影，因而也是在 $\mathscr{A}$ 上的投影. 所以又可记做 $\hat{z} = E(z|\mathscr{W}) = E(z|\mathscr{A})$.

2. $\hat{z} \in \mathscr{W}$ 为 z 的最小方差估计的充分必要条件是：对任意 $y \in \mathscr{W}$，必有 $E(z-\hat{z})y = 0$. 首先注意，由性质 1，任意 $y \in \mathscr{W}$ 可表 $y = \sum_{j=M}^{N} \beta_j a_j$，且 $\sum_{j=M}^{N} \beta_j^2 < \infty$. 所以 $Ey^2 = \sigma_a^2 \sum_{j=M}^{N} \beta_j^2 < \infty$. 下面用反证法来证明条件的必要性. 设 $\hat{z}$ 是 z 的最小方差估计，而且存在 $y \in \mathscr{W}$，使得 $E(z-\hat{z})y \neq 0$. 那么显然成立 $0 < Ey^2 < \infty$. 由最小方差性质，对任意实数 β 有

$$E(z-\hat{z})^2 \leqslant E(z-\hat{z}-\beta y)^2$$
$$= E(z-\hat{z})^2 + \beta^2 Ey^2 - 2\beta E(z-\hat{z})y$$

由此有 $\beta^2 Ey^2 \geqslant 2\beta E(z-\hat{z})y$. 取 $\beta = \frac{E(z-\hat{z})y}{Ey^2} \neq 0$，便导出 $1 \geqslant 2$. 从这一矛盾就证明了必要性. 再来证条件的充分性. 设 $\hat{z} \in \mathscr{W}$ 满足条件：$E(z-\hat{z})y = 0$，对任意 $y \in \mathscr{W}$ 成立. 那么，对任意 $z^* \in \mathscr{W}$，因为 $\hat{z} - z^* \in \mathscr{W}$，所以有 $E(z-\hat{z})(\hat{z} - z^*) = 0$. 于是

$$E(z-z^*)^2 = E(z-\hat{z}+\hat{z}-z^*)^2$$
$$= E(z-\hat{z})^2 + E(\hat{z}-z^*)^2 \geqslant E(z-\hat{z})^2$$

这就证明了 $\hat{z}$ 的最小方差性质。又可注意，上式的不等号当且仅当 $z^* = \hat{z}$ 时才能变成等号，所以 $\hat{z}$ 是唯一的。

3. z 的最小方差估计 $\hat{z}$ 存在而且唯一. 唯一性已经在性质 2 的充分性证明过程中得到. 现在来证明存在性. 令

$$\hat{z} = \frac{1}{\sigma_a^2} \sum_{j=M}^{N} (Eza_j)a_j.$$

注意 a_j 为白噪声，容易算出

$$0 \leqslant E(z - \hat{z})^2 = Ez^2 - 2Ez\hat{z} + E\hat{z}^2$$
$$= Ez^2 - \sum_{j=M}^{N} (Eza_j)^2/\sigma_a^2$$

所以 $\sum_{j=M}^{N} (Eza_j/\sigma_a^2)^2 \leqslant Ez^2/\sigma_a^2 < \infty$. 由性质 1，$\hat{z} \in \mathscr{A} = \mathscr{W}$. 又对任意 $y = \sum_{j=M}^{N} \beta_j a_j \in \mathscr{W}$，

$$E(z - \hat{z})y = \sum_{j=M}^{N} \beta_j(Eza_j) - \sum_{j=M}^{N} \beta_j(Eza_j) = 0$$

因此由性质 2，$\hat{z}$ 即为 z 的最小方差估计.

4. 若 $z = \beta_1 z_1 + \beta_2 z_2$，其中 β_1, β_2 是实数，则 $\hat{z} = \beta_1\hat{z}_1 + \beta_2\hat{z}_2$；若 $z \in \mathscr{W}$，则 $\hat{z} = z$；若 z 与 $\mathscr{W}$ 的元都独立，则 $\hat{z} = 0$；若 $w_k(M - 1 < k < N + 1)$ 是相互独立的随机序列，则任意 z 的最小方差估计为

$$\hat{z} = \sum_{j=M}^{N} (Ezw_j/Ew_j^2)w_j \tag{1.5.13}$$

这几条都很容易从性质 2 得到验证.

5. 若 $\boldsymbol{z}$ 是 r 维随机向量，且 M, N 有穷，则 $\boldsymbol{z}$ 的最小方差估计 $\hat{\boldsymbol{z}}$（即 $\hat{\boldsymbol{z}}$ 的各分量为 $\boldsymbol{z}$ 的对应分量的最小方差估计）可表为

$$\hat{\boldsymbol{z}} = (E\boldsymbol{z}\boldsymbol{w}^\tau)(E\boldsymbol{w}\boldsymbol{w}^\tau)^{-1}\boldsymbol{w} \tag{1.5.14}$$

其中 $\boldsymbol{w}^\tau = (w_M, w_{M+1}, \cdots, w_N)$. 这也容易由性质 2 验证.

三、参数估计的优效性概念

以上我们介绍了两种估计，这两种估计在概念上有本质的差

别．最小方差估计是用随机变量或序列的样本对另一随机变量作出估计，它们之间的概率分布是已知的，而最小二乘估计则是用随机变量或序列的样本去估计某些未知参数．前者主要在第七章用于解决随机序列的预报问题；后者主要用在第四章，以解决绪论中所说的“参数化”非线性估计问题． 除了最小二乘法以外，还可以用极大似然法或近似极大似然法来解决参数估计问题，这些方法我们将在今后用到它们的时候再作叙述．

从数理统计的角度怎样衡量参数估计的优劣程度，是另一个很重要的问题，我们在这里引进几个有关的定义．所谓参数估计，无非是根据某种原则，将随机序列的样本(即量测值) $y_1, y_2, \cdots, y_N$ 进行各种运算，从而对于未知参数向量 $\boldsymbol{\beta}$ 作出估计或判断．因此，一般可以把 $\boldsymbol{\beta}$ 的估计量 $\hat{\boldsymbol{\beta}}$ 表成样本的函数形式，即 $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}(y_1, y_2, \cdots, y_N)$，为了衡量 $\hat{\boldsymbol{\beta}}$ 与真值 $\boldsymbol{\beta}$ 的近似程度，需要以下一些概念：

1. 无偏性与渐近无偏性． 若估计量 $\hat{\boldsymbol{\beta}}$ 满足 $E\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}$，我们称 $\hat{\boldsymbol{\beta}}$ 为 $\boldsymbol{\beta}$ 的无偏估计． 若估计量只能满足 $\lim\limits_{N\to\infty} E\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}$，则称 $\hat{\boldsymbol{\beta}}$ 为 $\boldsymbol{\beta}$ 的渐近无偏估计．在很多情况下，寻找无偏估计很困难，但是容易找到渐近无偏估计．

2. 相容性． 若样本长度 $N\to\infty$ 时，估计量 $\hat{\boldsymbol{\beta}}$ 依概率收敛于 $\boldsymbol{\beta}$，即对任意小的 $\varepsilon > 0$，

$$\lim_{N\to\infty} P\{\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\| > \varepsilon\} = 0$$

其中 $\|\boldsymbol{\beta}\| = \max\{|\beta_1|, |\beta_2|, \cdots, |\beta_r|\}$ 是 $\boldsymbol{\beta}$ 的范数(或称模量)，这时我们称 $\hat{\boldsymbol{\beta}}$ 为 $\boldsymbol{\beta}$ 的相容估计．

3. 优效性与渐近优效性． 在相当一般的条件限制下，特别当限于讨论正态随机序列时，经典统计中的 Cramér-Rao 不等式（见[10])仍成立，即有

$$E(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^{\tau} \geq J_N^{-1}(\boldsymbol{\beta}) = \left(E\frac{\partial \log p}{\partial \boldsymbol{\beta}} \frac{\partial \log p}{\partial \boldsymbol{\beta}^{\tau}}\right)^{-1} \tag{1.5.15}$$

其中 p 是 $y_1, y_2, \cdots, y_N$ 的联合概率密度，$J_N^{-1}(\boldsymbol{\beta})$ 称为 Cramér-Rao 下界，这一公式我们将在附录 §4 中证明．若估计量 $\hat{\boldsymbol{\beta}}$ 能使 (1.5.15)式的等号成立，即

$$E(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})^{\tau}=J_N^{-1}(\boldsymbol{\beta}) \tag{1.5.16}$$

则称 $\hat{\boldsymbol{\beta}}$ 为优效估计．若估计量 $\hat{\boldsymbol{\beta}}$ 只能成立极限关系式

$$\lim_{N\to\infty} J_N^{\frac{1}{2}}(\boldsymbol{\beta})\,E(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})^{\tau}J_N^{\frac{1}{2}}(\boldsymbol{\beta})=I \tag{1.5.17}$$

其中 I 表示单位矩阵，则称 $\hat{\boldsymbol{\beta}}$ 为渐近优效估计．

4. 渐近正态性．若存在一个矩阵列 B_N，当 $N\to\infty$ 时，$B_NB_N^{\tau}$ 的主对角线都无限地增大，而且使得 $B_N(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})$ 的联合分布 F_N 收敛于正态分布 $\mathcal{N}(0, I)$，则称 $\hat{\boldsymbol{\beta}}$ 具有渐近正态性，简单地用 $B_N(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})\sim\mathcal{N}(0,I)$ 表示．

5. 优效渐近正态性．若 $\hat{\boldsymbol{\beta}}$ 具有渐近正态性，而且其中的 B_N 满足

$$\lim_{N\to\infty} B_NJ_N^{-1}(\boldsymbol{\beta})B_N^{\tau}=I \tag{1.5.18}$$

则称 $\hat{\boldsymbol{\beta}}$ 具有优效渐近正态性，即 $J_N^{\frac{1}{2}}(\boldsymbol{\beta})(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})\sim\mathcal{N}(0,I)$．

以上诸定义虽然都是从经典统计理论中搬过来的概念，但是，须要作两点说明．

(1) 如前已指出的，随机序列的参数估计与经典统计有一点本质性的差别．在经典统计中，样本 $y_1, y_2, \cdots, y_N$ 通常是相互独立同分布的随机变量，而参数 $\boldsymbol{\beta}$ 只是这一相同的分布中所含的未知参数（如正态分布的均值、方差等）． 在我们这里，$y_1, y_2, \cdots, y_N$ 是随机序列的一段样本，它们一般不是相互独立的，而参数 $\boldsymbol{\beta}$ 是这些随机变量的联合分布中的未知参数（如自回归序列的系数）．

(2) 估计量的渐近优效性和优效渐近正态性的渐近法则是不相同的．前者(即定义 3) 要求 $E(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})^{\tau}$ 与 $J_N^{-1}(\boldsymbol{\beta})$ 渐近相等，由于 $J_N^{-1}(\boldsymbol{\beta})$ 是估计误差 $(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})$ 的方差阵的下界，因此，渐近优效性又可以叫做渐近最小方差性． 而后者(即定义 5) 只是要求 $J_N^{\frac{1}{2}}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})$ 的分布 F_N 与正态分布 $N(0, I)$ 渐近相等，

或者说（$\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}$）的分布与 $N(0, J_N^{-1}(\boldsymbol{\beta}))$渐近相等，这里并不要求（$\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}$）的方差阵的收敛性．在一些数理统计著作中（例如[10]），曾经将这两种不同的渐近性质混淆起来，因而对某些估计类只证明了具有后一性质，却误认为也证明了前者．在本书中，为了澄清这种混乱，第一次采用了"优效渐近正态"一词，并且对某些估计量，分别证明了上述两种性质．有关的详细讨论见第四章§4．

最后还须指出，在具体使用参数估计方法时，我们总希望估计量能具有上述的各种优效性质，也就是说，这些性质是检验估计优劣的重要标准．不仅如此，这些性质在实际问题中还有重要的使用价值．例如，如果估计具有优效性或者渐近优效性，我们就知道估计的误差方差几乎达到了极小程度，根据 Cramér-Rao 下界还可以具体估算出其误差的方差大小(见第四、五章)．又如，当估计具有渐近正态性时，我们可以由此建立各种检验手段，以判断关于模型的假定的合理性，在第四、五、六、七诸章都要用到 $\hat{\boldsymbol{\beta}}$ 的渐近正态性质．其它优效性质也都是比较有用的．

第二章　线性模型及自相关与偏相关函数

§1. 随机线性模型

在绪论中，我们已经提到了随机差分方程(4)，在第一章里又给出了随机差分方程(1.3.24)的确切定义，本章将详细讨论形如绪论中(4)或(1.3.24)这样的随机线性模型．在此以前先通过两个例子简单说明随机序列、随机模型与时间序列应用之间的关系．

例 1. 在某一专用计算机的固定程序中，包括如下的简单迭代计算

$$x_t = \varphi_1 x_{t-1} \tag{2.1.1}$$

其中 φ_1 是固定的常数 ($|\varphi_1| < 1$)．由于计算机的字长有限，每次计算 (2.1.1) 式时都会有舍入误差．若以 x_t^0 和 x_t 分别表示计算机的计算值和真实值，那么二者之差 $w_t = x_t^0 - x_t$ 便是一个误差序列，我们现在来分析它的变化规律． 设在计算 x_t^0 时产生的舍入误差为 a_t，于是计算值 x_t^0 为

$$x_t^0 = \varphi_1 x_{t-1}^0 + a_t \tag{2.1.2}$$

其中 $\varphi_1 x_{t-1}^0$ 这一项表示 φ_1 与 x_{t-1}^0 的真实乘积值，由于机器有舍入误差 a_t，所以 x_t^0 有(2.1.2)式的形式． 联合 (2.1.1) 与 (2.1.2) 便可得到

$$w_t = \varphi_1 w_{t-1} + a_t \tag{2.1.3}$$

经验表明，舍入误差 a_t 近似为均匀分布的白噪声，其方差依计算机的字长而定．(2.1.3) 式就是计算 (2.1.1) 式时，计算误差序列所满足的随机模型，它是 (1.3.24) 式的一种特殊情况． 不过以后我们将主要讨论 a_t 为正态分布的情况．(2.1.3) 式和普通差分方程不同，由于 a_t 是随机序列，所以 w_t 也是随机序列．在每次使用程序 (2.1.1) 式时，由于初始值 x_0 不同，x_t 序列也各不相同，因而 a_t 序列也各不相同，于是 w_t 序列也各次取不同值，但是每次和它

相应的 a_t 都满足 (2.1.3)式. 如果用 B 表示一步延迟算子,即

$$Bw_t = w_{t-1} \tag{2.1.4}$$

那么 (2.1.3) 式可以用图 2.1.1 示意.

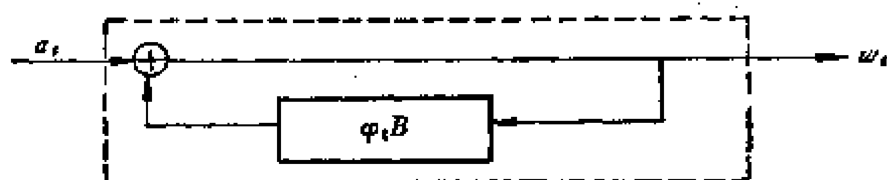


图 2.1.1 (2.1.3)式的系统示意框图

若把图 (2.1.1) 中用虚线包围的部分看做是一个线性系统的框图, a_t 为其输入, w_t 为其输出, 它恰好是工程技术人员所熟悉的一级反馈系统.

例 2. 空中飞行目标(如飞机、导弹或卫星等). 在某一空域飞行时,其加速度常常被视为随机过程, 在离散采样时, 就是随机序列. 比如在绪论的例 2 中我们曾把

$$u_t = \dot{x}_t - \dot{x}_{t-1}$$

看做是满足 (1.3.24) 式 (即绪论中(4)式) 的平稳序列, 并希望用时序分析方法估计 u_t 的模型参数. 这时, 我们不能象例 1 那样用简单推导列出 u_t 的模型.

通过以上两例和以前举过的很多例子, 我们来概括一下随机序列与随机模型的关系. 象例 1 那样从计算机和程序的构造与设计,能够准确导出误差序列所满足的随机模型,我们称它为能用物理方法列出的随机模型. 在这种情况下, 就不存在使用时序分析的问题了. 在实际中这种情况是少见的. 象例 2 那样, 对于物理过程 u_t, 并无物理方法能准确列出它的模型. 事实上, 我们说 u_t 具有 (1.3.24) 式的模型,如在第一章所说过的那样,只是一种近似地描述随机序列的手段, 即用具有有理谱的平稳序列来近似描述 u_t. 这时我们只能用 u_t 的样本序列来估计模型 (1.3.24) 式, 这就要用到时序分析方法. 这种情况在实际应用中是大量的, 如绪论中的例 1 至例 4 都属于这种情况.

有理谱与随机模型(1.3.24)式的关系已在第一章中说明了，这里再进一步归纳如下：因为一，具有有理谱(1.3.18)式的平稳序列必满足(1.3.24)式；二，(1.3.24)式的平稳解便于在最小方差意义下进行最佳预报或控制设计；三，有理谱能较好地逼近各种连续谱密度．所以，我们力图用(1.3.24)式模型来近似描述真实随机序列的"模型"．虽然(1.3.24)式模型的解不完全合于真实的随机序列，但由于我们根据真实随机序列的样本值，用了适当的估计方法求得(1.3.24)式的阶数和系数，在一定意义下能保证它们有相近的协方差函数．因而在解决最佳预报或控制时，便可使用模型(1.3.24)式以代替真实随机序列．

下面开始对模型(1.3.24)式进行较仔细的数学分析．先将(1.3.24)式重写于下，

$$\begin{aligned} & w_t - \varphi_1 w_{t-1} - \cdots - \varphi_p w_{t-p} \\ & = a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q} \end{aligned} \tag{2.1.5}$$

以后，我们总假定(2.1.5)式中的 a_t 为正态平稳白噪声，其方差 $Ea_t^2 = \sigma_a^2$，而且假定 $Ew_t a_s = 0$ $(t < s)$．用 B^k 表示 k 步延迟算子，即

$$B^k w_t = w_{t-k} \qquad B^k a_t = a_{t-k}$$

并令

$$\begin{aligned} \varphi(B) &= 1 - \varphi_1 B - \varphi_2 B^2 - \cdots - \varphi_p B^p \\ \theta(B) &= 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q \end{aligned} \tag{2.1.6}$$

则(2.1.5)式可以缩写成

$$\varphi(B) w_t = \theta(B) a_t \tag{2.1.7}$$

我们还总假设 $\varphi(B)$ 与 $\theta(B)$ 无公共因子．(2.1.5)式或(2.1.7)式的一种特殊情况是

$$w_t - \varphi_1 w_{t-1} - \cdots - \varphi_p w_{t-p} = a_t \tag{2.1.8}$$

或者写成

$$\varphi(B) w_t = a_t \tag{2.1.9}$$

另一种特殊情况是

$$w_t = a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q} \tag{2.1.10}$$

或者写成

$$w_t = \theta(B)a_t \tag{2.1.11}$$

如果 $z_t = w_t + \mu$，即 $Ez_t = \mu$，而且 w_t 满足（2.1.5）式，则有

$$\varphi(B)(z_t - \mu) = \theta(B)a_t \tag{2.1.12}$$

若记 $\varphi(B)\mu = (1 - \varphi_1 - \varphi_2 - \cdots - \varphi_p)\mu = C$，于是上式可以写成

$$\varphi(B)z_t = C + \theta(B)a_t \tag{2.1.13}$$

这就是随机序列 z_t 的均值不为零时的模型．这种模型不是我们要讨论的主要对象．因为 z_t 的相关函数与 w_t 的完全相同，所以只要讨论模型（2.1.5）式就够了．

虽然(2.1.9)，(2.1.11)式仅仅是(2.1.7)式的特别情况，但是，它们各自具有特殊的属性，这些属性以后经常用到．因此，我们对这三种情况分别给以不同的名称，这就是下面关于模型的分类：

1．可逆滑动平均模型．如果(2.1.11)式中的系数多项式 $\theta(\omega) = 0$ 的根全在单位圆外，即其根的模都大于1，我们就称(2.1.11)式为可逆滑动平均模型，其解 w_t 叫做可逆滑动平均序列．为了简便，我们又分别称它们为 MA（Moving Average）模型和 MA 序列．q 称为滑动平均阶数，$\theta_1, \theta_2, \cdots, \theta_q$ 和 σ_a^2 称为它们的参数．以后常用向量记号

$$\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_q)^\tau$$

还用缩写记号 MA$(0, q)$ 模型(或序列)，表示模型(或序列)是 q 阶纯滑动平均的．

2．平稳自回归模型．如果模型(2.1.8)或(2.1.9)式中的系数多项式 $\varphi(\omega) = 0$ 的根全在单位圆外，即其根的模都大于1，我们就称(2.1.8)式或(2.1.9)式为平稳自回归模型，其平稳解 w_t 叫做平稳自回归序列，为了简便，我们分别称它们为 AR（Autoregressive）模型和 AR 序列．p 称为自回归阶数，$\varphi_1, \varphi_2, \cdots, \varphi_p$ 和 σ_a^2 称为它们的参数．以后常用向量记号

$$\boldsymbol{\varphi} = (\varphi_1, \varphi_2, \cdots, \varphi_p)^\tau$$

还用缩写记号 AR$(p, 0)$ 模型(或序列)，表示模型(或序列)是

p 阶纯自回归的.

3. 平稳自回归-可逆滑动平均混合模型. 如果模型 (2.1.5) 式或 (2.1.7) 式中的系数多项式 $\varphi(\omega)$ 和 $\theta(\omega)$ 无公共因子,而且分别满足上面的平稳性条件和可逆性条件，我们就称这一模型为平稳自回归-可逆滑动平均混合模型，其平稳解叫做平稳自回归-可逆滑动平均混合序列. 并分别简称为 ARMA 模型和 ARMA 序列. 用 (p, q) 表示其阶数,前者表示自回归的阶数，后者表示滑动平均的阶数. $\boldsymbol{\varphi}$, $\boldsymbol{\theta}$ 和 σ_a^2 为其参数,为了方便,有时用

$$\boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\varphi} \\ \boldsymbol{\theta} \end{pmatrix} \tag{2.1.14}$$

表示两组系数组成的参数向量;用

$$\boldsymbol{\alpha} = \begin{pmatrix} \boldsymbol{\beta} \\ \sigma_a^2 \end{pmatrix} \tag{2.1.15}$$

表示全部参数组成的向量. 同样也采用 ARMA(p, q) 模型(或序列)的缩写记号. 由此分类可以看出，ARMA$(p, 0)$ 模型就是 AR$(p, 0)$ 模型; ARMA$(0, q)$ 模型就是 MA$(0, q)$ 模型. 平稳自回归模型和可逆滑动平均模型都是混合模型的特殊情况.

上面给了模型的平稳性和可逆性的定义. 根据这些定义，如果给了一个随机模型 (2.1.5) 式 (或 (2.1.7) 式),我们可以求出相应的方程 $\varphi(\omega) = 0$ 和 $\theta(\omega) = 0$ 的根,检验这些根是否全在单位圆外,以此来判断模型是否为平稳的和可逆的. 另外,也可以根据代数方程的根和系数的关系，把平稳性和可逆性条件转化成关于参数 $\boldsymbol{\varphi}$ 和 $\boldsymbol{\theta}$ 的约束条件，这样会更方便些. 为此我们再引入以下两个概念.

1. 平稳域. 设模型 (2.1.5) 式的自回归阶数为 p，凡是使 $\varphi(\omega) = 0$ 的根全在单位圆外的系数向量 $\boldsymbol{\varphi}$，构成一个 p 维实向量空间的子集，记做 $\Phi^{(p)}$，$\Phi^{(p)} = \{\boldsymbol{\varphi}: \boldsymbol{\varphi}^{\tau} = (\varphi_1, \varphi_2, \cdots, \varphi_p)$，$\varphi(\omega) = 0$ 的根全在单位圆外$\}$，$\Phi^{(p)}$ 称为模型的平稳域. 模型 (2.1.5) 式为平稳的充分必要条件是 $\boldsymbol{\varphi}$ 属于 $\Phi^{(p)}$.

2. 可逆域. 设模型 (2.1.5) 式的滑动平均阶数为 q，凡是使

$\theta(\omega)=0$ 的根全在单位圆外的系数向量 θ，构成一个 q 维实向量空间的子集，记做 $\Theta^{(q)}=\{\theta:\theta^{\tau}=(\theta_1,\theta_2,\cdots,\theta_q),\Theta(\omega)=0$ 的根全在单位圆外$\}$. $\Theta^{(q)}$ 称为模型的可逆域. 模型 (2.1.5) 式为可逆的充分必要条件是 θ 属于 $\Theta^{(q)}$.

例 3. AR(1,0) 模型(即 $p=1,q=0$)

$$w_t-\varphi_1 w_{t-1}=a_t$$

其相应的代数方程为 $\varphi(\omega)=1-\varphi_1\omega=0$，其根为 $\omega_1=\dfrac{1}{\varphi_1}$; 为使 $|\omega_1|=|\varphi_1^{-1}|>1$，必须且只须 $|\varphi_1|<1$. 所以 $\Phi^{(1)}=\{\varphi_1:-1<\varphi_1<1\}$ 就是模型 AR(1, 0) 或 ARMA(1, q) 的平稳域.

例 4. AR(2, 0) 模型

$$w_t-\varphi_1 w_{t-1}-\varphi_2 w_{t-2}=a_t$$

相应的代数方程为 $\varphi(\omega)=1-\varphi_1\omega-\varphi_2\omega^2=0$，它的两个根为

$$\omega_1=\frac{-\varphi_1-\sqrt{\varphi_1^2+4\varphi_2}}{2\varphi_2},\quad \omega_2=\frac{-\varphi_1+\sqrt{\varphi_1^2+4\varphi_2}}{2\varphi_2}$$

而且 $\omega_1\omega_2=-\dfrac{1}{\varphi_2}$, $\quad\omega_1+\omega_2=-\dfrac{\varphi_1}{\varphi_2}$, 因而有

$$\varphi_2+\varphi_1=-\frac{1}{\omega_1\omega_2}\pm\left(\frac{1}{\omega_1}+\frac{1}{\omega_2}\right)=1-\left(1\mp\frac{1}{\omega_1}\right)\left(1\mp\frac{1}{\omega_2}\right)$$

注意当 ω_1 为复数时，由于 φ_1,φ_2 是实数，ω_2 必为 ω_1 的共轭复数；当 ω_1 为实数时，ω_2 也必是实数. 于是，如果 $|\omega_i|>1(i=1,2)$，那么 $|\varphi_2|<1$，而且无论 ω_i 为实数还是复数，都有

$$\varphi_2\pm\varphi_1=1-\left(1\mp\frac{1}{\omega_1}\right)\left(1\mp\frac{1}{\omega_2}\right)<1$$

反之，如果 $|\varphi_2|<1$ 且 $\varphi_2\pm\varphi_1<1$ $\left(\text{即}\left(1\mp\dfrac{1}{\omega_1}\right)\left(1\mp\dfrac{1}{\omega_2}\right)>0\right)$，那么从前者推出 $|\omega_i|(i=1,2)$ 中至少有一个大于1，不妨设 $|\omega_1|>1$. 当 ω_1 为复数时，必有 $\omega_2=\bar{\omega}_1$，这时 $|\omega_2|=|\omega_1|>1$；当 ω_1,ω_2 都为实数时，由于 $|\omega_1|>1$，意味着 $1\mp\dfrac{1}{\omega_1}>0$，所以从

$$\left(1\mp\frac{1}{\omega_1}\right)\left(1\mp\frac{1}{\omega_2}\right)>0$$

推得 $1\pm\frac{1}{\omega_2}>0$，即 $-1<\frac{1}{\omega_2}<1$ 或 $|\omega_2|>1$. 这就证明了无论 ω_i 为实数或复数，都有 $|\omega_i|>1$. 综合上面的证明，为使 $|\omega_i|>1$，其充分必要条件是 (φ_1,φ_2) 属于

$$\Phi^{(2)}=\{\boldsymbol{\varphi}:\boldsymbol{\varphi}=(\varphi_1,\varphi_2)^\tau,\ -1<\varphi_2<1,\ \varphi_2+\varphi_1<1,\\ \varphi_2-\varphi_1<1\}$$

这就是 AR(2, 0) 或 ARMA(2, q) 的平稳域，图 2.1.2(a) 给出了这个区域.

例 5. MA(0, 1) 模型

与 AR(1, 0) 模型类似：$\Theta^{(1)}=\{\theta:-1<\theta<1\}$.

例 6. MA(0, 2) 模型

与 AR(2, 0) 模型类似：$\Theta^{(2)}=\{\boldsymbol{\theta}:\ \boldsymbol{\theta}=(\theta_1,\theta_2)^\tau,\ \theta_2+\theta_1<1,\ \theta_2-\theta_1<1,\ -1<\theta_2<1\}$，图 2.1.2(b) 给出了这个区域.

例 7. ARMA(1, 1) 模型

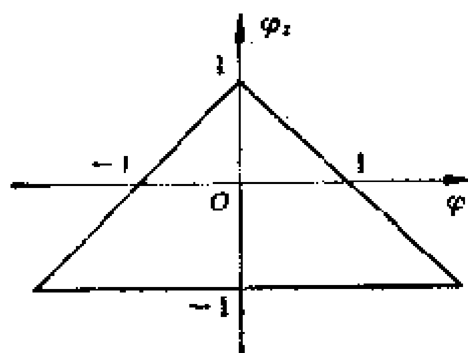
它的平稳可逆域是 $(\Phi^{(1)},\Theta^{(1)})=\{(\varphi_1,\theta_1):\ -1<\varphi_1<1,\ -1<\theta_1<1\}$，图 2.1.2(c) 给出了这个区域.

当 ARMA(p, q) 模型的 p, q 不超过 2 时，由以上几例可以组合出它的平稳可逆域；当 p, q 超过 2 以后，其平稳可逆域都比较复杂.

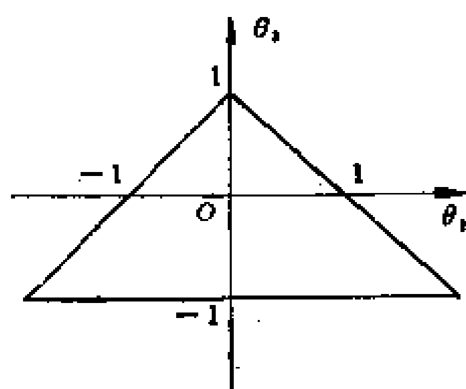
高阶模型参数的平稳性与可逆性检验.

由上述知，当 p 和 q 不超过 2 时，模型 ARMA(p, q) 的参数平稳域与可逆域可以具体给出，而当 p 和 q 超过 2 时，模型的平稳可逆域就变得比较复杂，不可能用简单的关系式表示. 但是，可以有办法检验它们的参数是否属于平稳可逆域. 当然，最直接的办法是求解代数方程，检查它的根是否在单位圆外，但这种方法往往计算量较大，况且有时并不需要了解特征方程根的具体数值，这里我们将介绍两种比较简单的判别准则.

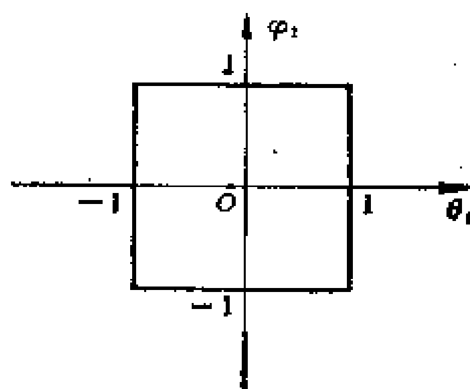
定义. 一个系统称为稳定的当且仅当系统特征多项式



(a) AR(2,0) 的平稳域



(b) MA(0,2) 的可逆域



(c) ARMA(1,1) 的平稳可逆域

图 2.1.2 低阶模型的平稳域与可逆域

$$F(z) = a_n z^n + a_{n-1} z^{n-1} + \cdots + a_1 z + a_0, \quad a_n > 0$$

的根在复平面的单位圆内.

1. Schur-Cohn 准则

令 $F(z)$ 的 S-C 行列式为

$$\Delta_k = \begin{vmatrix} a_0 & & & & a_n & a_{n-1} & \cdots & a_{n-k+1} \\ a_1 & a_0 & & & & \ddots & & \vdots \\ \vdots & & \ddots & & & & \ddots & \vdots \\ a_{k-1} & \cdots & & a_0 & & & & a_n \\ \bar{a}_n & & & & \bar{a}_0 & \bar{a}_1 & \cdots & \bar{a}_{k-1} \\ \bar{a}_{n-1} & \bar{a}_n & & & & \ddots & & \\ \vdots & & \ddots & & & & \ddots & \\ a_{n-k+1} & \cdots & & \bar{a}_n & & & & \bar{a}_0 \end{vmatrix}, \quad k = 1, 2, \cdots, n$$

$\bar{a}_k$ 为 a_k 的共轭复数,且令 $\Delta_0 = -1$.

样本多项式 $F(z)$ 稳定的充分必要条件是：它的 $S\text{-}C$ 行列式满足

$$\Delta_k \begin{cases} <0, & \text{当 } k \text{ 是奇数} \\ >0, & \text{当 } k \text{ 是偶数} \end{cases}$$

如果上述条件满足则 $F(z)=0$ 的所有的根在单位圆内，否则至少有一个根不在单位圆内，但不能判断究竟有几个，也不能判断这些根是在圆外还是在圆上。

2. Jury 准则. Jury 检验准则依赖于以下的表格：

行	z^0	z^1	z^2	$\cdots$	z^{n-k}	$\cdots$	z^{n-1}	z^n
1	a_0	a_1	a_2	$\cdots$	a_{n-k}	$\cdots$	a_{n-1}	a_n
2	a_n	a_{n-1}	a_{n-2}	$\cdots$	a_n	$\cdots$	a_1	a_0
3	b_0	b_1	b_2	$\cdots$	b_{n-1-k}	$\cdots$	b_{n-1}	
4	b_{n-1}	b_{n-2}	b_{n-3}	$\cdots$	b_k	$\cdots$	b_0	
5	c_0	c_1	c_2	$\cdots$	$\cdots$	c_{n-2}		
6	c_{n-2}	c_{n-3}	c_{n-4}	$\cdots$	$\cdots$	c_0		
$\cdots\cdots$								
$2n-5$	p_0	p_1	p_2	p_3				
$2n-4$	p_3	p_2	p_1	p_0				
$2n-3$	q_0	q_1	q_2					

表中第一行中依次排列 $F(z)$ 的系数，第二行是第一行中的逆序，其余各行中的元素如下定义：

$$b_k = \begin{vmatrix} a_0 & a_{n-k} \\ a_n & a_k \end{vmatrix}, \quad c_k = \begin{vmatrix} b_0 & b_{n-1-k} \\ b_{n-1} & b_k \end{vmatrix}$$

$$\cdots\cdots$$

$$q_0 = \begin{vmatrix} p_0 & p_3 \\ p_3 & p_0 \end{vmatrix}, \quad q_1 = \begin{vmatrix} p_0 & p_2 \\ p_3 & p_1 \end{vmatrix}, \quad q_2 = \begin{vmatrix} p_0 & p_1 \\ p_3 & p_2 \end{vmatrix}$$

$F(z)=0$ 在单位圆外和圆上无根的充分必要条件是

$$F(1)>0, \quad F(-1)\begin{cases} >0 & \text{当 } n \text{ 是偶数} \\ <0 & \text{当 } n \text{ 是奇数} \end{cases}$$

以及

$|a_0| < |a_n|, |b_0| > |b_{n-1}|, |c_0| > |c_{n-2}| \cdots, |q_0| > |q_2|$
这 $n-1$ 个关系式都成立.

在具体检验 ARMA 模型的平稳性和可逆性时，要注意首先将 $\varphi(B)$ 和 $\theta(B)$ 化成规定的形式 $F(z)$，当然在 ARMA 中系数都是实的，然后才能作进一步讨论，例如要检验 $\varphi(B) = 1 - \varphi_1 B - \cdots \varphi_p B^p$ 的所有根是否在单位圆外，就应转化为检验

$$\bar{\varphi}(z) = z^p - \varphi_1 z^{p-1} - \cdots - \varphi_p$$

的根是否在单位圆内，其对应的系数序列就应该是

$$(-\varphi_p, -\varphi_{p-1}, \cdots -\varphi_1, 1)$$

同理，关于 $\theta(B)$ 的可逆域的讨论，涉及的系数序列是

$$(-\theta_q, -\theta_{q-1}, \cdots -\theta_1, 1)$$

有关这两个准则可参考[22].

在给出了以上的线性模型概念以后，我们来揭开第一章第二节所举的六个伪随机数列构造的秘密. 它们是分别按照以下的模型构造的:

$$w_k = a_k - 0.5a_{k-1} \qquad k = 1, 2, \cdots \tag{2.1.16a}$$

$$w_k = a_k - a_{k-1} + 0.24a_{k-2} \qquad k = 1, 2, \cdots \tag{2.1.16b}$$

$$w_k - 0.5w_{k-1} = a_k \qquad k = 1, 2, \cdots \tag{2.1.16c}$$

$$w_k - w_{k-1} + 0.24w_{k-2} = a_k \qquad k = 1, 2, \cdots \tag{2.1.16d}$$

$$w_k - 0.3w_{k-1} = a_k - 0.6a_{k-1} \qquad k = 1, 2, \cdots \tag{2.1.16e}$$

$$z_k - 1.5z_{k-1} + 0.5z_{k-2} = a_k \qquad k = 1,2,\cdots \tag{2.1.16f}$$

诸模型中的 a_k 是具有相同方差 $\sigma_a^2 = 1$ 的伪正态白噪声序列，其构造方法见[17]. 以上诸式中凡用到初始值 $w_0, w_{-1}, \cdots, z_0, z_{-1}, \cdots$ 或 $a_0, a_{-1}, \cdots$ 等均用零代之. 易见 (2.1.16a) 是 MA(0, 1) 模型，(2.1.16b) 是 MA(0, 2)，(2.1.16c) 是 AR(1, 0)，(2.1.16d) 是 AR(2, 0)，(2.1.16 e) 是 ARMA(1, 1)，但是，(2.1.16 f) 不是 AR(2, 0) 模型，因为 $\varphi(\omega) = 1 - 1.5\omega + 0.5\omega^2 = 0$ 的根为 $\omega_1 = 1, \omega_2 = 2$，其中 $|\omega_1| = 1 \ngtr 1$，它不满足平稳性条件. 由诸式可见，这些序列之间有着明显的差别，但是，由此得到的六个样本序列（见第一章 § 2 图 1.1.1）之间的差异却并不那么显著. 这

是因为随机序列的一段样本值只能部分地反映该随机序列的本质，而且只有在对样本值进行必要的分析运算后，才能揭示出这部分本质. 今后我们将逐步深入地介绍分析随机序列的样本值的方法. 为此，先要引进 w_k 与 a_k 相互关系的另外两种表达形式，即传递形式和逆转形式.

1. 传递形式 —— ARMA 模型平稳解的一种表达方式.

现在讨论 ARMA 模型 (2.1.5) 式的解. 若在 (2.1.5) 式中给了白噪声序列 a_t 的一串样本值 $\cdots a_{-1}, a_0, a_1, \cdots$，而 w_t 的相应的样本值可以用普通的差分方程解法来求解.

例如 AR(1, 0) 模型 $w_t - \varphi_1 w_{t-1} = a_t$，注意 $|\varphi_1| < 1$，可以解出

$$
\begin{aligned}
w_k &= \varphi_1 w_{k-1} + a_k = \varphi_1^2 w_{k-2} + \varphi_1 a_{k-1} + a_k = \cdots \\
&= \sum_{j=0}^{M-1} \varphi_1^j a_{k-j} + \varphi_1^M w_{k-M} = \lim_{M\to\infty} \sum_{j=0}^{M-1} \varphi_1^j a_{k-j} \\
&= \sum_{j=0}^{\infty} \varphi_1^j a_{k-j}
\end{aligned}
$$

其中，因为 w_k 为平稳序列，$Ew_k^2 = \gamma_0 < \infty$，所以当 $M \to \infty$ 时 $\varphi_1^k w_{k-M}$ 均方收敛于零. 若 $\sum_{j=0}^{\infty} \varphi_1^j B^j = (1 - \varphi_1 B)^{-1}$，则

$$w_k = (1 - \varphi_1 B)^{-1} a_k$$

由此可见，延迟算子 B 所组成的多项式，可以形式地进行代数运算. 但是必须注意，$(1 - \varphi_1 B)^{-1} a_k$ 的真实含义是 $\sum_{j=0}^{\infty} \varphi_1^j B^j a_k = \sum_{j=0}^{\infty} \varphi_1^j a_{k-j}$，因此，须在条件 $|\varphi_1| < 1$ 成立时无穷求和才有意义. 如果这一条件满足，在求 AR(1, 0) 模型的解 w_t 时，可以很方便地看成将

$$(1 - \varphi_1 B) w_t = a_t$$

两边用算子 $(1 - \varphi_1 B)^{-1}$ 加以作用而得到 $w_t = (1 - \varphi_1 B)^{-1} a_t$. 这种形式运算的方法今后将经常用到.

现在考虑一般 ARMA 模型 (2.1.7) 式的解，依上所述有

$$w_t = \varphi^{-1}(B)\theta(B)a_t = \theta(B)\varphi^{-1}(B)a_t = \Psi(B)a_t$$

$$= \sum_{k=0}^{\infty} \psi_k B^k a_t = \sum_{k=0}^{\infty} \psi_k a_{t-k} \qquad (2.1.17)$$

其中 $\Psi(B) = \sum_{k=0}^{\infty} \psi_k B^k = \varphi^{-1}(B)\theta(B)$. 系数 ψ_k 是由以下等式

$$\varphi(B)\Psi(B) = \theta(B)$$

比较两边 B^k 的系数而得到的. 易见 $\psi_0 = 1$, 其它的 $\psi_1, \psi_2 \cdots$ 可由递推形式逐个获得. 比如, 等式右边 B 的系数是 $(-\theta_1)$, 等式左边 B 的系数是 $(\psi_1 - \varphi_1)$ 所以由 $-\theta_1 = \psi_1 - \varphi_1$ 得到 $\psi_1 = \varphi_1 - \theta_1$. (2.1.17) 式就是 ARMA 模型(2.1.7)式的传递形式, 其平稳解 w_t 用白噪声 a_t 的无穷滑动和的形式表达出来. 为了保证(2.1.17)式中的无穷滑动和是在均方意义下收敛的, 须要证明 $\Sigma\psi_k^2 < \infty$. 事实上我们将证明一个更强的结论, 即当 $k \to \infty$ 时, $|\psi_k|$ 依负指数律趋于零. 在证明之前, 首先注意一个明显的事实: 对 MA $(0, q)$ 模型而言, (2.1.10) 式本身就是传递形式, 所以有

$$\psi_0 = 1; \ \psi_k = -\theta_k, \ 1 \leqslant k \leqslant q; \ \psi_k = 0, \ k > q \qquad (2.1.18)$$

对于 AR 或 ARMA 模型而言我们再看两个例子.

例 8. ARMA(1, 1) 模型 $(1 - \varphi_1 B)w_t = (1 - \theta_1 B)a_t$ 其传递形式为

$$w_t = (1 - \varphi_1 B)^{-1}(1 - \theta_1 B)a_t = \Psi(B)a_t$$

其中 $\Psi(B) = (1 - \varphi_1 B)^{-1}(1 - \theta_1 B)$, 即 $(1 - \varphi_1 B)\Psi(B) = 1 - \theta_1 B$. 比较两边 B^k 的系数容易求得 $\psi_0 = 1$ 和

$$\psi_k = \varphi_1^{k-1}(\varphi_1 - \theta_1), \ k \geqslant 1 \qquad (2.1.19)$$

另外, 直接展开 $\Psi(B) = (1 - \varphi_1 B)^{-1}(1 - \theta_1 B) = (1 - \theta_1 B) \sum_{k=0}^{\infty} \varphi_1^k B^k = 1 + \sum_{k=1}^{\infty} \varphi_1^{k-1}(\varphi_1 - \theta_1)B^k$, 也同样得到 (2.1.19) 式的结果.

例 9. AR(2, 0) 模型 $(1 - \varphi_1 B - \varphi_2 B^2)w_t = a_t$, 其传递形式为

$$w_t = (1 - \varphi_1 B - \varphi_2 B^2)^{-1}a_t = \Psi(B)a_t$$

其中 $\Psi(B)=(1-\varphi_1B-\varphi_2B^2)^{-1}$ 即 $(1-\varphi_1B-\varphi_2B^2)\Psi(B)=1$. 比较 B^k 的系数得到

$$\psi_0=1,\ \psi_1=\varphi_1,\ \psi_k=\varphi_1\psi_{k-1}+\varphi_2\psi_{k-2},\ k\geqslant 2 \tag{2.1.20}$$

对于一般 ARMA(p, q) 模型而言，给出 ψ_k 的通用表达式是比较麻烦的，而且我们也不需要，我们关心的是 ψ_k 的收敛情况，即当 $k\to\infty$ 时 ψ_k 趋于零的速度. 注意

$$\begin{aligned}1-\theta_1B-\cdots-\theta_qB^q&=\varphi(B)\Psi(B)\\&=\left(1-\sum_{j=1}^{p}\varphi_jB^j\right)\left(\sum_{k=0}^{\infty}\psi_kB^k\right)\\&=\sum_{k=0}^{\infty}\psi_kB^k-\sum_{k=0}^{\infty}\sum_{j=1}^{p}\varphi_j\psi_kB^{k+j}\\&=\sum_{k=0}^{\infty}\psi_kB^k-\sum_{l=1}^{\infty}\sum_{j=1}^{\min(p,l)}\varphi_j\psi_{l-j}B^l\\&=1-\sum_{l=1}^{\infty}\left(\psi_l-\sum_{j=1}^{\min(p,l)}\varphi_j\psi_{l-j}\right)B^l\end{aligned}$$

可见，当 $l>q$ 时

$$\psi_l-\sum_{j=1}^{\min(p,l)}\varphi_j\psi_{l-j}=0$$

特别当 $l>\max(p, q)$ 时，上式又可简记为

$$\varphi(B)\psi_l=0,\quad l>\max(p, q) \tag{2.1.21}$$

这说明当 $l>\max(p, q)$，ψ_l 满足差分方程 (2.1.21). 这是普通的常系数线性差分方程，而且易知 $\psi_0, \psi_1, \cdots, \psi_{\max(p,q)}$ 都是有穷值，根据差分方程理论[14]. 由于多项式 $\varphi(\omega)$ 满足平稳性条件(其根全在单位圆外)，所以存在正常数 g_1, g_2 使得

$$|\psi_l|\leqslant g_1e^{-g_2l},\quad l\geqslant 0 \tag{2.1.22}$$

即 $|\psi_l|$ 被一负指数函数所控制. 回忆一下 (1.5.10) 式，这正相当于那里所限定的情况.

2. 逆转形式——ARMA 模型中的白噪声 a_t 表达成 w_t 的形式.

下面讨论传递形式的相反问题，即 (2.1.7) 式的白噪声 a_t 如何用 w_t 表达出来．与上面的步骤完全类似，由 (2.1.7) 式得到

$$a_t = \theta^{-1}(B)\varphi(B)w_t = \pi(B)w_t = w_t - \sum_{j=1}^{\infty} \pi_j w_{t-j} \quad (2.1.23)$$

其中 $\pi(B)$ 满足 $\theta(B)\pi(B) = \varphi(B)$，稍有差异的是，前面我们令 $\Psi(B) = 1 + \sum_{j=1}^{\infty} \psi_j B^j$，而这里却令 $\pi(B) = 1 - \sum_{j=1}^{\infty} \pi_j B^j$，这种符号上的差异在讨论预报问题时各有其方便之处．(2.1.23) 式称为 (2.1.7) 式的逆转形式．对于 AR(p, 0) 模型而言，模型 (2.1.8) 式本身就是逆转形式，因此有

$$\pi_k = \varphi_k, \quad 1 \leqslant k \leqslant p, \quad \pi_k = 0, \quad k > p \quad (2.1.24)$$

对于一般 ARMA 模型 (2.1.7) 式而言，仍有类似 (2.1.22) 式的公式，而且不妨仍用同样的常数 g_1, g_2 (如若不同，可分别取较大者和较小者)，即

$$|\pi_j| \leqslant g_1 e^{-g_2 j}, \quad j \geqslant 0 \quad (2.1.25)$$

例 10. ARMA(1, 1) 模型 $(1 - \varphi_1 B)w_t = (1 - \theta_1 B)a_t$，其逆转形式为

$$\begin{aligned} a_t &= (1 - \theta_1 B)^{-1}(1 - \varphi_1 B)w_t \\ &= (1 - \varphi_1 B)\sum_{j=0}^{\infty} \theta_1^j B^j w_t = \left(1 - \sum_{j=1}^{\infty} \pi_j B^j\right)w_t \end{aligned}$$

与例 8 相类似可导出 $\pi_0 = 1$,

$$\pi_k = -\theta_1^{k-1}(\theta_1 - \varphi_1), \quad k \geqslant 1 \quad (2.1.26)$$

以上是关于随机模型的一些基本内容．下一节我们将讨论由 ARMA 模型解出的 ARMA 序列的基本性质．

§2. ARMA 序列的自相关与偏相关函数

本节介绍 ARMA 序列的自相关函数与偏相关函数及其性质，并讨论它们与模型参数之间的关系．根据第一章的论述，平稳序列为 ARMA 序列的必要和充分条件是它具有有理谱，现在，我们就利用有理谱平稳序列所满足的 ARMA 模型，导出其自相关函

数的性质和表达式.

利用 ARMA 序列的传递形式 (2.1.17) 式和白噪声性质可知，w_t 的自协方差函数为

$$\gamma_k = \gamma_k(w) = Ew_t w_{t+k} = \sigma_a^2 \sum_{j=0}^{\infty} \phi_j \phi_{k+j} = \gamma_{-k}, \quad k \geq 0 \tag{2.2.1}$$

在容易发生混淆时,我们才用 $\gamma_k(w)$ 表示，以标明它是 w_t 的自协方差函数. 对于 w_t 的自相关函数 ρ_k 也如此处理,这时

$$\rho_k = \rho_k(w) = \frac{Ew_t w_{t+k}}{\sqrt{Ew_t^2 Ew_{t+k}^2}}$$

$$= \left(\sum_{j=0}^{\infty} \phi_j \phi_{j+k}\right)\left(\sum_{j=0}^{\infty} \phi_j^2\right)^{-1} = \rho_{-k} \quad k \geq 0 \tag{2.2.2}$$

(2.2.1) 和 (2.2.2) 式只是 γ_k 和 ρ_k 的表达方式之一，特别是(2.2.1)式在第一章 §3 中已经介绍过了. 下面分别讨论 AR$(p, 0)$, MA$(0, q)$ 及 ARMA(p, q) 序列的自相关函数的性质，以及其它表达方式.

一、MA(0, q) 序列的自相关函数

当 w_t 为 MA$(0, q)$ 序列时,将 (2.1.8) 式代入 (2.2.1),(2.2.2) 式得

$$\gamma_k = \begin{cases} \sigma_a^2(1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2) & k = 0 \\ \sigma_a^2(-\theta_k + \theta_1\theta_{k+1} + \cdots + \theta_{q-k}\theta_q) & i \leq k \leq q \\ 0 & k > q \end{cases} \tag{2.2.3}$$

$$\rho_k = \begin{cases} 1 & k = 0 \\ \dfrac{-\theta_k + \theta_1\theta_{k+1} + \cdots + \theta_{q-k}\theta_q}{1 + \theta_1^2 + \cdots + \theta_q^2} & i \leq k \leq q \\ 0 & k > q \end{cases} \tag{2.2.4}$$

这两式说明,当 w_t 与 w_τ 的相隔步数 $|t - \tau| > q$ 时，w_t 与 w_τ 不相关,即 MA 序列的自协方差（或自相关）函数 γ_k（或 ρ_k），从 $k > q$ 以后全部为零. 我们称这一性质为"截尾". 反之，可以证

明，若一平稳序列其自协方差（或自相关）函数为截尾，那么它必是 MA(0,q) 序列（证明见附录§1）.

公式(2.2.3)和(2.2.4)式已经表明了 γ_k 或 ρ_k 与 MA(0, q) 模型参数 θ, σ_a^2 的相互关系，根据这些关系，由 γ_k 或 $\rho_k(0\leqslant k\leqslant q)$ 也可以反解出 θ 与 σ_a^2.

二、AR(p,0) 序列的自相关函数

AR(p, 0) 序列的自相关函数没有 MA(0, q) 序列那样的截尾性质. 我们将(2.1.8)式写成

$$w_t=\varphi_1 w_{t-1}+\varphi_2 w_{t-2}+\cdots+\varphi_p w_{t-p}+a_t \qquad (2.2.5)$$

以 $w_{t-k}(k>0)$ 乘以上式两边，再求各自的均值得

$$\gamma_k=\varphi_1\gamma_{k-1}+\varphi_2\gamma_{k-2}+\cdots+\varphi_p\gamma_{k-p},\quad k>0 \qquad (2.2.6)$$

因此又有

$$\rho_k=\varphi_1\rho_{k-1}+\varphi_2\rho_{k-2}+\cdots+\varphi_p\rho_{k-p},\quad k>0 \qquad (2.2.7)$$

或者再将上式写成

$$\varphi(B)\rho_k=0,\qquad k>0 \qquad (2.2.8)$$

这就是 AR(p, 0) 的自相关函数所满足的差分方程，与(2.1.21)式相同，只是初始值不同而已. 因此，ρ_k 也满足(2.1.22)式. 当 $\varphi(\omega)=0$ 的根都不相同时，设 $\omega_1, \omega_2, \cdots, \omega_p$ 为它的 p 个根，根据差分方程理论，在初值 $\rho_0, \rho_{-1}, \cdots, \rho_{-p+1}$ 给定下(2.2.8)式的通解为

$$\rho_k=c_1\omega_1^{-k}+c_2\omega_2^{-k}+\cdots+c_p\omega_p^{-k}\qquad k>-p \qquad (2.2.9)$$

其中的系数 $c_1, c_2, \cdots, c_p$ 由 p 个初始值 $\rho_0,\rho_{-1},\cdots,\rho_{-p+1}$ 代入(2.2.9)式而定. 但是初值的选定不能任意. 根据 ρ_k 的对称性，$\rho_k=\rho_{-k}$，我们分别取 $k=0, 1, 2, \cdots, p-1$，由此及(2.2.9)式可得如下的 p 阶线性方程组

$$\begin{cases}c_1+c_2+\cdots+c_p=1\\ c_1(\omega_1^k-\omega_1^{-k})+c_2(\omega_2^k-\omega_2^{-k})+\cdots+c_p(\omega_p^k-\omega_p^{-k})=0\end{cases}$$

$$k=1, 2, \cdots, p-1 \qquad (2.2.10)$$

由此即可以定出常数 $c_1, c_2, \cdots, c_p$. 当 $\varphi(\omega)=0$ 有重根时，

则(2.2.9)式稍有不同，比如 $\omega_1=\omega_2=\cdots=\omega_r$，那么(2.2.9)式的前 r 项应为 $(c_1+c_2k+\cdots+c_rk^{r-1})\omega_1^{-k}$，其它重根也类似，在此不详细论述．总之，无论 $\varphi(\omega)=0$ 有无重根，序列 ρ_k 都不是截尾的序列，但是却能满足(2.1.22)式．我们称这样的 ρ_k 为“拖尾”的；即 ρ_k 的尾部不全为零，但又被负指数函数控制．

例 1. AR(2,0) 序列的自相关函数应满足方程

$$\rho_k=\varphi_1\rho_{k-1}+\varphi_2\rho_{k-2},\quad k>0$$

当 $1-\varphi_1\omega-\varphi_2\omega^2=0$ 的根 $\omega_1\neq\omega_2$ 时，其解 ρ_k 为

$$\rho_k=c_1\omega_1^{-k}+c_2\omega_2^{-k},\quad k\geqslant-1$$

$$\begin{cases}c_1+c_2=1\\ c_1(\omega_1-\omega_1^{-1})+c_2(\omega_2-\omega_2^{-1})=0\end{cases}$$

由此定出

$$c_1=\frac{\omega_2-\omega_2^{-1}}{(\omega_2-\omega_2^{-1})-(\omega_1-\omega_1^{-1})}$$

$$c_2=\frac{\omega_1-\omega_1^{-1}}{(\omega_1-\omega_1^{-1})-(\omega_2-\omega_2^{-1})}$$

当 $\omega_1=\omega_2$ 时

$$\rho_k=(c_1+c_2k)\omega_1^{-k},\qquad k\geqslant-1$$

由 $\rho_0=1$ 可知 $c_1=1$；又根据 $\rho_1=\rho_{-1}$ 可得

$$(c_1+c_2)\omega_1^{-1}=(c_1-c_2)\omega_1$$

由此有 $1+c_2=(1-c_2)\omega_1^2$，从而解出 $c_2=\dfrac{\omega_1^2-1}{1+\omega_1^2}$．由于 $|\omega_1|>1$，$|\omega_2|>1$，因此无论 ω_1,ω_2 是否相等，ρ_k 总被负指数函数所控制，即满足(2.1.22)式．

上面介绍了如何从 $\varphi_1,\varphi_2,\cdots,\varphi_p$ 求 ρ_k 的方法．反之，也可以从 ρ_k 求出 $\varphi_1,\varphi_2,\cdots,\varphi_p$．由(2.2.7)式，取 $k=1,2,\cdots,p$，可得下面的方程组：

$$\begin{cases}\rho_1=\varphi_1+\varphi_2\rho_1+\cdots+\varphi_p\rho_{p-1}\\ \rho_2=\varphi_1\rho_1+\varphi_2+\cdots+\varphi_p\rho_{p-2}\\ \cdots\cdots\cdots\\ \rho_p=\varphi_1\rho_{p-1}+\varphi_2\rho_{p-2}+\cdots+\varphi_p\end{cases}\tag{2.2.11}$$

由此清楚地看出参数向量 φ 是线性方程组(2.2.11)式的解．这个方程组的系数全部由 $\rho_1, \rho_2, \cdots, \rho_p$ 组成，我们称它为 Yule-Walker 方程，它在今后是非常有用的．最后为求参数 σ_a^2、将 (2.1.8) 式两边平方后再求均值，得到

$$\begin{aligned}\sigma_a^2 &= Ea_t^2 = E(w_t - \varphi_1 w_{t-1} - \cdots - \varphi_p w_{t-p})^2 \\ &= \gamma_0 - 2\sum_{j=1}^{p}\varphi_j\gamma_j + \sum_{i,j=1}^{p}\varphi_i\varphi_j\gamma_{j-i}\end{aligned}$$

注意由 (2.2.11) 式可得

$$\sum_{j=1}^{p}\varphi_j\rho_j = \sum_{i,j=1}^{p}\varphi_i\varphi_j\rho_{j-i}$$

由此又有

$$\sum_{j=1}^{p}\varphi_j\gamma_j = \sum_{i,j=1}^{p}\varphi_i\varphi_j\gamma_{j-i} \tag{2.2.12}$$

以此代入 σ_a^2 得

$$\sigma_a^2 = \gamma_0 - \sum_{j=1}^{p}\varphi_j\gamma_j = \gamma_0 - \sum_{i,j=1}^{p}\varphi_i\varphi_j\gamma_{j-i} \tag{2.2.13}$$

三、ARMA(p, q) 序列的自相关函数

如果 w_t 是 ARMA(p, q) 序列，那么它的自相关函数比 AR 序列还要复杂一些．我们先求出 w_t 与 a_t 的互相关函数 $\gamma_k(w, a)$．利用 w_t 的传递形式 (2.1.17) 式可得

$$\begin{aligned}\gamma_k(w, a) &= Ew_l a_{l+k} = E\left(\sum_{j=0}^{\infty}\psi_j a_{l-j} a_{l+k}\right) \\ &= \sum_{j=0}^{\infty}\psi_j E(a_{l-j}a_{l+k}) = \begin{cases}\sigma_a^2\psi_{-k} & k \leqslant 0 \\ 0 & k > 0\end{cases}\end{aligned} \tag{2.2.14}$$

把 (2.1.5) 式两边乘以 w_{t-k} 再求均值，又得到

$$\begin{aligned}&\gamma_k - \varphi_1\gamma_{k-1} - \cdots - \varphi_p\gamma_{k-p} \\ &= \gamma_k(w, a) - \theta_1\gamma_{k-1}(w, a) - \cdots - \theta_q\gamma_{k-q}(w, a)\end{aligned} \tag{2.2.15}$$

将 (2.2.14) 式代入 (2.2.15) 式，并将等式两边除以 γ_0 可得

$$\varphi(B)\rho_k = 0, \qquad k > q \tag{2.2.16}$$

于是又一次出现了类似于(2.1.21)式的方程，这里k是从$q+1$开始的．关于它的解的形式已经讨论过了，不同之处在于确定初始值的方法．仅就$\varphi(\omega)=0$无重根的情况说明之，这时知道了初始值$\rho_q, \rho_{q-1}, \cdots, \rho_{q-p+1}$，通解就可写为：

$$\rho_k = c_1\omega_1^{-k} + c_2\omega_2^{-k} + \cdots + c_p\omega_p^{-k}, \quad k > q-p \tag{2.2.17}$$

由于$\gamma_k = \gamma_{-k}$，所以$\gamma_0, \gamma_1, \cdots, \gamma_p$可由(2.2.14)式和(2.2.15)式解出，如果$q > p$，$\gamma_{p+1}, \gamma_{p+2}, \cdots, \gamma_q$又可由(2.2.15)式逐个递推求出，于是初值$\rho_q, \rho_{q-1}, \cdots, \rho_{q-p+1}$也便不难得到了．同样$\rho_1, \cdots, \rho_{q-p}$也不难得到．

反之，若已知$\rho_0, \rho_1, \rho_2, \cdots$，利用(2.2.16)式取$k=q+1, q+2, \cdots, q+p$可得$p$阶线性方程组

$$\begin{cases} \rho_{q+1} = \varphi_1\rho_q + \varphi_2\rho_{q-1} + \cdots + \varphi_p\rho_{q-p+1} \\ \rho_{q+2} = \varphi_1\rho_{q+1} + \varphi_2\rho_q + \cdots + \varphi_p\rho_{q-p+2} \\ \cdots\cdots \\ \rho_{q+p} = \varphi_1\rho_{q+p-1} + \varphi_2\rho_{q+p-2} + \cdots + \varphi_p\rho_q \end{cases} \tag{2.2.18}$$

由它即可解出自回归参数向量$\boldsymbol{\varphi}$来．再令$\bar{w}_k = w_k - \varphi_1 w_{k-1} - \varphi_2 w_{k-2} - \cdots - \varphi_p w_{k-\varphi}$，则$\bar{w}_t$是以$\theta_1, \theta_2, \cdots, \theta_q$为参数的MA$(0, q)$序列，而$\bar{w}_t$的自协方差函数$\gamma_k(\bar{w})$可由$\gamma_k(w)$计算出来，即

$$\begin{aligned} \gamma_k(\bar{w}) &= E\Big(w_t - \sum_{j=1}^{p}\varphi_j w_{t-j}\Big)\Big(w_{t+k} - \sum_{i=1}^{p}\varphi_i w_{t+k-i}\Big) \\ &= \gamma_k - \sum_{i=1}^{p}\varphi_i\gamma_{k-i} - \sum_{j=1}^{p}\varphi_j\gamma_{k+j} + \sum_{i,j=1}^{p}\varphi_i\varphi_j\gamma_{k+j-i} \\ &= \sum_{i,j=0}^{p}\varphi_i\varphi_j\gamma_{k+j-i} \end{aligned} \tag{2.2.19}$$

在上式最后的求和项中，我们把φ_0取做-1．有了$\gamma_k(\bar{w})$，我们就可以将它代人(2.2.3)式解出滑动平均参数向量$\boldsymbol{\theta}$和σ_a^2．

现在我们归纳一下ARMA序列的自相关函数ρ_k的性质．当

$p=0$ 时，序列退化为 MA 型，ρ_k 是截尾的；当 $q=0$ 时，序列退化为 AR 型，ρ_k 是拖尾的，它们能用 (2.2.9) 式（当 $\varphi(\omega)=0$ 无重根时）统一表达；当 $q\neq 0$，$p\neq 0$ 时，ρ_k 也是拖尾的，而且在 $k>q-p$ 后，ρ_k 能用 (2.2.9) 式（当 $\varphi(\omega)=0$ 无重根时）统一表达，但是 $\rho_0,\rho_1,\cdots,\rho_{q-p}$（当 $q\geqslant p$ 时）这几个 ρ_k 不能用 (2.2.9) 式统一表示．这些性质在以后将要用到．

利用上面关于 ARMA 序列自协方差函数的知识，我们将第一章所举的前五个伪随机序列的自协方差函数一一求出来．

模型 (2.1.16 c) 的自相关函数 $\rho_k=(0.5)^k$.

模型 (2.1.16 d) 的自相关函数 $\rho_k=c_1\omega_1^{k}+c_2\omega_2^{-k}$，由于 $\varphi(\omega)=1-\omega+0.24\omega^2=0$ 的根是 $\omega_1=(0.4)^{-1}$，$\omega_2=(0.6)^{-1}$，所以

$$\rho_k=c_1(0.4)^k+c_2(0.6)^k$$

再根据本节例 1 可知 $\rho_k=\dfrac{63}{31}(0.6)^k-\dfrac{32}{31}(0.4)^k$.

模型 (2.1.16 a) 的自相关函数

$$\rho_1=\frac{-0.5}{1+(0.5)^2}=\frac{-0.5}{1.25}=-0.4,\quad \rho_k=0,\quad k\geqslant 2$$

模型 (2.1.16 b) 的自相关函数

$$\rho_1=\frac{-1-0.24}{1+1+(0.24)^2}=\frac{-1.24}{2.0576}=-\frac{775}{1286}$$

$$\rho_2=\frac{0.24}{2.0576}=\frac{75}{643},\quad \rho_k=0,\quad k\geqslant 3$$

模型 (2.1.16 e) 的自协方差函数

$$\gamma_k=c_1(0.3)^k,\qquad k>q-p=0$$

而 γ_0 不能由上式表达．为了定出系数 c_1，需要由 (2.2.14) 式和 (2.2.15) 式计算初值 γ_1，先由 (2.2.14) 式得

$$\gamma_0(w,a)=\sigma_a^2\psi_0=1$$

$$\gamma_1(w,a)=\sigma_a^2\psi_1=0.3-0.6=-0.3$$

再代入 (2.2.15) 式，分别取 $k=0$ 和 1 得

$$\gamma_0-0.3\gamma_{-1}=1.18,\qquad \gamma_1-0.3\gamma_0=-0.6$$

注意 $\gamma_{-1}=\gamma_1$，由以上二式即可解出 $\gamma_1=-\frac{246}{910}$，$\gamma_0=\frac{100}{91}$.
因此，$c_1=-\frac{82}{91}$，于是 $\rho_k=-0.82\times(0.3)^k$.

通过这几个典型的数值例子，我们可以熟悉本节的内容．此外，以这五个自相关函数还可以看出，它们的统计性质是有明显差别的．分别从这些序列的一段样本作出的样本自协方差函数（见(1.2.1)式和图 3.2.1），是上述理论自协方差函数的估计值，所以理应比原始的样本（见第一章§2 图 1.2.1）更加突出了各序列的本质差别；至于它们能够反映到怎样的程度，将在第三章中论述．

四、偏相关函数

依上所述，对于平稳序列而言，其自相关函数的截尾性是 MA 序列所特有的标志．但是拖尾性则是 AR 和 ARMA 序列所共有的特征．那么，AR 序列有没有独自的特征呢？为了回答这个问题，我们引入偏相关函数的概念．

对于 $k=1,2,\cdots$ 我们分别来考虑用 $w_{t-k},w_{t-k+1},\cdots,w_{t-1}$ 对 w_t 作最小方差估计（见第一章§5），即选择系数 $\varphi_{k1},\varphi_{k2},\cdots,\varphi_{kk}$，使得

$$\begin{aligned}\delta&=E\Big(w_t-\sum_{j=1}^{k}\varphi_{kj}w_{t-j}\Big)^2\\&=\gamma_0-2\sum_{j=1}^{k}\varphi_{kj}\gamma_j+\sum_{i,j=1}^{k}\varphi_{kj}\varphi_{ki}\gamma_{j-i}\end{aligned}\tag{2.2.20}$$

达到极小值．分别对 $\varphi_{kj}(j=1,2,\cdots,k)$ 求 δ 的偏导数 $\frac{\partial\delta}{\partial\varphi_{kj}}$，并令其为零，即得到 φ_{kj} 所应满足的方程

$$\begin{pmatrix}1&\rho_1&\cdots&\rho_{k-1}\\\rho_1&1&\cdots&\rho_{k-2}\\&\cdots\cdots&&\\\rho_{k-1}&\rho_{k-2}&\cdots&1\end{pmatrix}\begin{pmatrix}\varphi_{k1}\\\varphi_{k2}\\\vdots\\\varphi_{kk}\end{pmatrix}=\begin{pmatrix}\rho_1\\\rho_2\\\vdots\\\rho_k\end{pmatrix}\tag{2.2.21}$$

我们称序列 $\varphi_{kk}(k\geqslant1)$ 为 w_t 的偏相关函数．

当 w_t 为 AR$(p,0)$ 序列时，先不必解(2.2.21)式，由(2.2.20)式容易看出，对于 $k \geq p$，可从 δ 的极小性直接定出 $\varphi_{kj}(1 \leq j \leq k)$. 事实上，若 $k \geq p$，则以(2.1.8)式代入(2.2.20)式得到

$$\delta = E\left(\sum_{j=1}^{p} \varphi_j w_{t-j} - \sum_{j=1}^{k} \varphi_{kj} w_{t-j} + a_t\right)^2$$

$$= E\left(a_t + \sum_{j=1}^{p} (\varphi_j - \varphi_{kj}) w_{t-j} - \sum_{j=p+1}^{k} \varphi_{kj} w_{t-j}\right)^2$$

$$= \sigma_a^2 + E\left(\sum_{j=1}^{p} (\varphi_j - \varphi_{kj}) w_{t-j} - \sum_{j=p+1}^{k} \varphi_{kj} w_{t-j}\right)^2 \geq \sigma_a^2$$

显而易见，为使 δ 达到极小值$(=\sigma_a^2)$，应当取

$$\varphi_{kj} = \begin{cases} \varphi_j, & 1 \leq j \leq p \\ 0, & p+1 \leq j \leq k \end{cases} \qquad k \geq p \qquad (2.2.22)$$

由此可见，AR$(p, 0)$ 序列的偏相关函数 φ_{kk}，在 $k > p$ 以后全等于零，即偏相关函数是截尾的. 在附录§1中还将证明，偏相关函数的截尾性也是平稳序列为 AR 型的充分条件，也就是说，它是 AR 序列所特有的标志.

下面来讨论 φ_{kj} 的递推算法和偏相关函数的概率意义，以及 ARMA 或 MA 序列的 φ_{kk} 的拖尾性. 首先，由(2.2.21)式解出的 φ_{kj} 有以下递推形式：

$$\begin{cases} \varphi_{11} = \rho_1 \\ \varphi_{k+1,k+1} = \left(\rho_{k+1} - \sum_{j=1}^{k} \rho_{k+1-j}\varphi_{kj}\right)\left(1 - \sum_{j=1}^{k} \rho_j \varphi_{kj}\right)^{-1} \\ \varphi_{k+1,j} = \varphi_{kj} - \varphi_{k+1,k+1}\varphi_{k,k-(j-1)} \\ \qquad j = 1, 2, \cdots, k \end{cases} \qquad (2.2.23)$$

为证此，令

$$\xi_k = \begin{pmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{pmatrix}, \quad \tilde{\xi}_k = \begin{pmatrix} \rho_k \\ \rho_{k-1} \\ \vdots \\ \rho_1 \end{pmatrix}, \quad \eta_k = \begin{pmatrix} \varphi_{k1} \\ \varphi_{k2} \\ \vdots \\ \varphi_{kk} \end{pmatrix}, \quad \zeta_k = \begin{pmatrix} \varphi_{k1} \\ \varphi_{k2} \\ \vdots \\ \varphi_{k,k-1} \end{pmatrix}$$

$$R_k=\begin{pmatrix}1 & \rho_1 & \cdots & \rho_{k-1}\\ \rho_1 & 1 & \cdots & \rho_{k-2}\\ & \cdots\cdots & & \\ \rho_{k-1} & \rho_{k-2} & \cdots & 1\end{pmatrix},\quad T=\begin{pmatrix} & & & 1\\ 0 & & 1 & \\ & \cdot^{\cdot^{\cdot}} & & \\ 1 & & & 0\end{pmatrix}$$

易见 $T^{-1}=T$，$TR_kT=R_k$，$T\boldsymbol{\xi}_k=\bar{\boldsymbol{\xi}}_k$，于是 (2.2.21) 式对 $k+1$ 而言可以写成

$$R_{k+1}\boldsymbol{\eta}_{k+1}=\boldsymbol{\xi}_{k+1}$$

或

$$\begin{pmatrix}R_k & \bar{\boldsymbol{\xi}}_k\\ \bar{\boldsymbol{\xi}}_k^{\tau} & 1\end{pmatrix}\begin{pmatrix}\boldsymbol{\zeta}_{k+1}\\ \varphi_{k+1,k+1}\end{pmatrix}=\begin{pmatrix}\boldsymbol{\xi}_k\\ \rho_{k+1}\end{pmatrix}$$

由分块矩阵运算可得

$$R_k\boldsymbol{\zeta}_{k+1}+\bar{\boldsymbol{\xi}}_k\varphi_{k+1,k+1}=\boldsymbol{\xi}_k,\quad \bar{\boldsymbol{\xi}}_k^{\tau}\boldsymbol{\zeta}_{k+1}+\varphi_{k+1,k+1}=\rho_{k+1}$$

由第一式，

$$\begin{aligned}\boldsymbol{\zeta}_{k+1}&=R_k^{-1}\boldsymbol{\xi}_k-\varphi_{k+1,k+1}R_k^{-1}\bar{\boldsymbol{\xi}}_k=\boldsymbol{\eta}_k-\varphi_{k+1,k+1}TR_k^{-1}\boldsymbol{\xi}_k\\ &=\boldsymbol{\eta}_k-\varphi_{k+1,k+1}T\boldsymbol{\eta}_k\end{aligned}\tag{2.2.24}$$

以此代入第二式得

$$\begin{aligned}\rho_{k+1}&=\varphi_{k+1,k+1}+\bar{\boldsymbol{\xi}}_k^{\tau}(\boldsymbol{\eta}_k-\varphi_{k+1,k+1}T\boldsymbol{\eta}_k)\\ &=\varphi_{k+1,k+1}-\varphi_{k+1,k+1}\boldsymbol{\xi}_k^{\tau}\boldsymbol{\eta}_k+\bar{\boldsymbol{\xi}}_k^{\tau}\boldsymbol{\eta}_k\end{aligned}$$

由此即解出

$$\varphi_{k+1,k+1}=(\rho_{k+1}-\bar{\boldsymbol{\xi}}_k^{\tau}\boldsymbol{\eta}_k)(1-\boldsymbol{\xi}_k^{\tau}\boldsymbol{\eta}_k)^{-1}\tag{2.2.25}$$

(2.2.24) 和 (2.2.25) 式就是 (2.2.23) 式的缩写形式. $\varphi_{11}=\rho_1$ 是显然的.

其次，我们指出偏相关函数的概率意义. 利用最小方差估计的记号，考虑 $\hat{w}_t=E(w_t|w_{t-1},w_{t-2},\cdots,w_{t-k+1})$ 和 $\hat{w}_{t-k}^*=E(w_{t-k}|w_{t-1},w_{t-2},\cdots,w_{t-k+1})$，并令 $\tilde{w}_t\equiv w_t-\hat{w}_t$，$\tilde{w}_t^*=w_{t-k}-\hat{w}_{t-k}^*$ 以及

$$\psi_{kk}\equiv(E\tilde{w}_t\tilde{w}_t^*)(E\tilde{w}_t^2E\tilde{w}_t^{*2})^{-\frac{1}{2}}\tag{2.2.26}$$

这是 $\tilde{w}_t$ 和 $\tilde{w}_t^*$ 的相关系数，或者称为在给定 $w_{t-1},w_{t-2},\cdots,w_{t-k+1}$ 的条件下，w_t 与 w_{t-k} 的条件相关系数. 再令

$$\boldsymbol{x}=(w_t,w_{t-k})^{\tau},\qquad \boldsymbol{y}=(w_{t-1},w_{t-2},\cdots,w_{t-k+1})^{\tau}$$

于是根据最小方差估计的性质(1.5.14)式有

$$\hat{\boldsymbol{x}} = E(\boldsymbol{x}|\boldsymbol{y}) = (E\boldsymbol{x}\boldsymbol{y}^{\tau})(E\boldsymbol{y}\boldsymbol{y}^{\tau})^{-1}\boldsymbol{y}$$

$$E\tilde{\boldsymbol{x}}\tilde{\boldsymbol{x}}^{\tau} = E(\boldsymbol{x}-\hat{\boldsymbol{x}})(\boldsymbol{x}-\hat{\boldsymbol{x}})^{\tau}$$
$$= E\boldsymbol{x}\boldsymbol{x}^{\tau} - E\boldsymbol{x}\hat{\boldsymbol{x}}^{\tau} - E\hat{\boldsymbol{x}}\boldsymbol{x}^{\tau} + E\hat{\boldsymbol{x}}\hat{\boldsymbol{x}}^{\tau}$$

注意 $E(\boldsymbol{x}-\hat{\boldsymbol{x}})\hat{\boldsymbol{x}}^{\tau}=0$, 即 $E\boldsymbol{x}\hat{\boldsymbol{x}}^{\tau}=E\hat{\boldsymbol{x}}\hat{\boldsymbol{x}}^{\tau}$, 所以

$$E\tilde{\boldsymbol{x}}\tilde{\boldsymbol{x}}^{\tau} = E\boldsymbol{x}\boldsymbol{x}^{\tau} - E\hat{\boldsymbol{x}}\hat{\boldsymbol{x}}^{\tau}$$
$$= E\boldsymbol{x}\boldsymbol{x}^{\tau} - (E\boldsymbol{x}\boldsymbol{y}^{\tau})(E\boldsymbol{y}\boldsymbol{y}^{\tau})^{-1}(E\boldsymbol{y}\boldsymbol{x}^{\tau})$$

仍然使用前边的记号 $\boldsymbol{\xi}_k$, $\boldsymbol{\eta}_k$, R_k 等,再注意

$$E\boldsymbol{x}\boldsymbol{x}^{\tau} = \gamma_0\begin{pmatrix} 1 & \rho_k \\ \rho_k & 1 \end{pmatrix}$$

$$E\boldsymbol{x}\boldsymbol{y}^{\tau} = \gamma_0\begin{pmatrix} \rho_1 & \rho_2 & \cdots & \rho_{k-1} \\ \rho_{k-1} & \rho_{k-2} & \cdots & \rho_1 \end{pmatrix} = \gamma_0\begin{pmatrix} \boldsymbol{\xi}_{k-1}^{\tau} \\ \bar{\boldsymbol{\xi}}_{k-1}^{\tau} \end{pmatrix}$$

$$E\boldsymbol{y}\boldsymbol{y}^{\tau} = \gamma_0 R_{k-1}$$

故有

$$E\tilde{\boldsymbol{x}}\tilde{\boldsymbol{x}}^{\tau} = \begin{pmatrix} E\tilde{w}_t^2 & E\tilde{w}_t\tilde{w}_t^* \\ E\tilde{w}_t^*\tilde{w}_t & E\tilde{w}_t^{*2} \end{pmatrix}$$
$$= \gamma_0\left\{\begin{pmatrix} 1 & \rho_k \\ \rho_k & 1 \end{pmatrix} - \begin{pmatrix} \boldsymbol{\xi}_{k-1}^{\tau} \\ \bar{\boldsymbol{\xi}}_{k-1}^{\tau} \end{pmatrix} R_{k-1}^{-1}(\boldsymbol{\xi}_{k-1}, \bar{\boldsymbol{\xi}}_{k-1})\right\}$$
$$= \gamma_0\left\{\begin{pmatrix} 1 & \rho_k \\ \rho_k & 1 \end{pmatrix} - \begin{pmatrix} \boldsymbol{\xi}_{k-1}^{\tau} \\ \bar{\boldsymbol{\xi}}_{k-1}^{\tau} \end{pmatrix} (R_{k-1}^{-1}\boldsymbol{\xi}_{k-1}, R_{k-1}^{-1}\bar{\boldsymbol{\xi}}_{k-1})\right\}$$
$$= \gamma_0\left\{\begin{pmatrix} 1 & \rho_k \\ \rho_k & 1 \end{pmatrix} - \begin{pmatrix} \boldsymbol{\xi}_{k-1}^{\tau} \\ \bar{\boldsymbol{\xi}}_{k-1}^{\tau} \end{pmatrix} (\boldsymbol{\eta}_{k-1}, T\boldsymbol{\eta}_{k-1})\right\}$$
$$= \gamma_0\left\{\begin{pmatrix} 1 & \rho_k \\ \rho_k & 1 \end{pmatrix} - \begin{pmatrix} \boldsymbol{\xi}_{k-1}^{\tau}\boldsymbol{\eta}_{k-1} & \boldsymbol{\xi}_{k-1}^{\tau}T\boldsymbol{\eta}_{k-1} \\ \boldsymbol{\xi}_{k-1}^{\tau}T\boldsymbol{\eta}_{k-1} & \boldsymbol{\xi}_{k-1}^{\tau}\boldsymbol{\eta}_{k-1} \end{pmatrix}\right\}$$

由此和(2.2.25)式可知

$$\psi_{kk} = (E\tilde{w}_t\tilde{w}_t^*)(E\tilde{w}_t^2 E\tilde{w}_t^{*2})^{-\frac{1}{2}}$$
$$= (\rho_k - \boldsymbol{\xi}_{k-1}^{\tau}T\boldsymbol{\eta}_{k-1})(1-\boldsymbol{\xi}_{k-1}^{\tau}\boldsymbol{\eta}_{k-1})^{-1} = \varphi_{kk}$$

这说明,从其概率意义而言,偏相关系数 φ_{kk}是在给定 $w_{t-1}, w_{t-2}, \cdots, w_{t-k+1}$ 的条件下 w_t 与 w_{t-k} 的条件相关系数. 更确切地说,是 w_t 与 w_{t-k} 扣去它们关于 $w_{t-1}, w_{t-2}, \cdots, w_{t-k+1}$ 的最小方差

估计后的残差的相关系数. 由这一性质，我们将不难证明偏相关函数 φ_{kk} 被负指数函数所控制.

现在令

$$u_k=\sum_{j=1}^{k-1}\pi_j w_{t-j},\qquad v_k=\sum_{j=k}^{\infty}\pi_j w_{t-j}$$

根据 w_t 的逆转形式(2.1.23)式和最小方差估计的性质，

$$\begin{aligned} w_t &= a_t+u_k+v_k \\ \tilde{w}_t &= w_t-\hat{w}_t = a_t+u_k+v_k \\ &\quad - E(a_t+u_k+v_k|w_{t-1},w_{t-2},\cdots,w_{t-k+1}) \\ &= a_t+u_k+v_k-u_k-\hat{v}_k = a_t+\tilde{v}_k \end{aligned}$$

由此知 $E\tilde{w}_t^2=Ea_t^2+E\tilde{v}_t^2\geqslant\sigma_a^2$. 又由(2.1.25)式，

$$Ev_k^2=\sum_{i,j=k}^{\infty}\pi_j\gamma_{j-i}\pi_i\leqslant\gamma_0 g_1^2\sum_{i,j=0}^{\infty}e^{-2g_2k}e^{-g_2(i+j)}\leqslant g_3e^{-g_4k}$$

式中 g_3 和 g_4 是适当的正常数. 再由 a_t 与 $w_{t-j}(j>0)$ 不相关可得 $E\tilde{w}_t\tilde{w}_t^*=E(a_t+\tilde{v}_k)\tilde{w}_t^*=E\tilde{v}_k\tilde{w}_t^*$，因此由 Schwarz 不等式，

$$\begin{aligned} |E\tilde{w}_t\tilde{w}_t^*| &\leqslant (E\tilde{v}_k^2E\tilde{w}_t^{*2})^{\frac{1}{2}}\leqslant(Ev_k^2Ew_{t-k}^2)^{\frac{1}{2}} \\ &\leqslant\sqrt{\gamma_0}\cdot\sqrt{g_3e^{-\frac{1}{2}g_4k}}=g_5e^{-g_6k} \end{aligned}$$

于是最后得到

$$\begin{aligned} |\varphi_{kk}| &= |\phi_{kk}|=|(E\tilde{w}_t\tilde{w}_t^*)(E\tilde{w}_t^2)^{-1}| \\ &\leqslant\frac{g_5e^{-g_6k}}{\sigma_a^2}=g_7e^{-g_8k} \end{aligned}$$

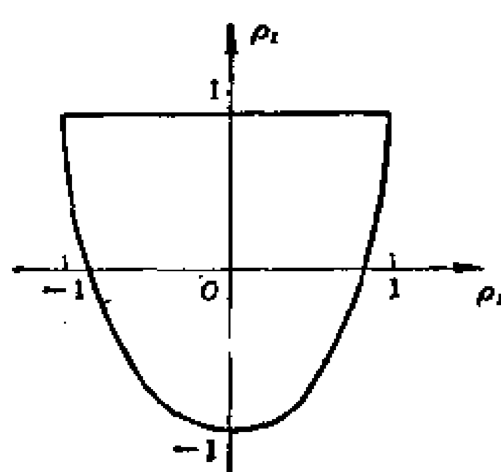
这就是(2.1.22)式的形式. 所以对于 ARMA 或 MA 序列 φ_{kk} 是拖尾的.

五、ARMA 序列的自相关函数的允许域

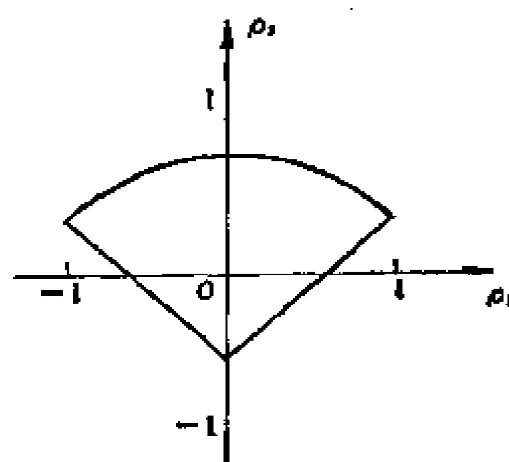
在第一章里我们讲过，只要序列 γ_k 满足对称性和非负定性，它就一定是某一平稳序列的自协方差函数. 但是，为要序列 γ_k (或 ρ_k) 成为 ARMA 序列的自协方差函数(或自相关函数)，除了需要满足对称性和非负定性以外，还须与 γ_k 相应的谱密度具有

(1.3.18)式的形式．换言之，由 γ_k 所决定的模型(2.1.5)式要具有平稳性和可逆性．这样在 p, q 固定以后，ARMA(p, q) 序列的自协方差函数（或自相关函数）所能允许的取值范围叫做它的允许域．一个自相关函数列 ρ_k，仅当它在此域内时，它才是 ARMA(p, q) 序列的自相关函数．因此，允许域可以被用来检验一串 ρ_k 是否为 ARMA(p, q) 的相关函数列．对于一般的允许域问题，这里不仔细讨论．

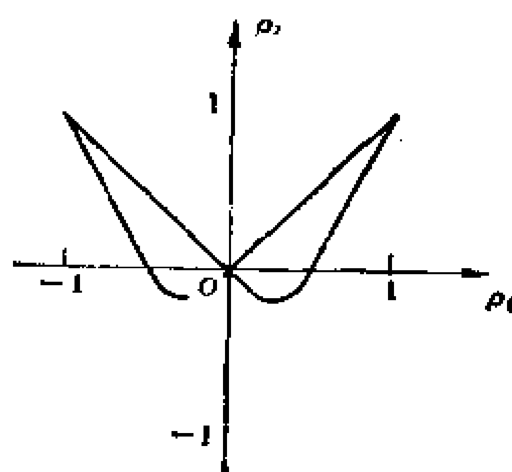
最后，为了熟悉上述概念，我们再举几个例子．在这些例子中给出序列的自相关与偏相关函数和允许域．并在图 2.2.1 和图 2.2.2 中给出它们的示意图．



(a) AR(2,0) 的允许域



(b) MA(0,2) 的允许域



(c) ARMA(1,1) 的允许域

图 2.2.1　低阶模型的允许域

例 2. MA(0, 1) 序列 $w_t = a_t - \theta_1 a_{t-1}$

$$\gamma_0 = (1+\theta_1^2)\sigma_a^2, \quad \gamma_1 = -\theta_1\sigma_a^2, \quad \gamma_k = 0, \quad k \geq 2$$

$$\rho_0 = 1, \qquad \rho_1 = \frac{-\theta_1}{1+\theta_1^2}, \quad \rho_k = 0, \quad k \geq 2$$

其 φ_{kj} 由下式决定

$$\begin{pmatrix} 1 & \frac{-\theta_1}{1+\theta_1^2} & & 0 \\ -\frac{\theta_1}{1+\theta_1^2} & 1 & \ddots & \\ & \ddots & \ddots & -\frac{\theta_1}{1+\theta_1^2} \\ 0 & & \frac{-\theta_1}{1+\theta_1^2} & 1 \end{pmatrix} \begin{pmatrix} \varphi_{k1} \\ \varphi_{k2} \\ \vdots \\ \varphi_{kk} \end{pmatrix} = \begin{pmatrix} \frac{-\theta_1}{1+\theta_1^2} \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

或者

$$\begin{pmatrix} 1+\theta_1^2 & -\theta_1 & & 0 \\ -\theta_1 & 1+\theta_1^2 & \ddots & \\ & \ddots & \ddots & -\theta_1 \\ 0 & & -\theta_1 & 1+\theta_1^2 \end{pmatrix} \begin{pmatrix} \varphi_{k1} \\ \varphi_{k2} \\ \vdots \\ \varphi_{kk} \end{pmatrix} = \begin{pmatrix} -\theta_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

令

$$\Delta_k = \begin{vmatrix} 1+\theta_1^2 & -\theta_1 & & 0 \\ -\theta_1 & \ddots & \ddots & -\theta_1 \\ 0 & & -\theta_1 & 1+\theta_1^2 \end{vmatrix}$$

那么由行列式展开可知 $\Delta_k = (1+\theta_1^2)\Delta_{k-1} - \theta_1^2\Delta_{k-2} (k \geq 3)$，这个差分方程的通解是

$$\Delta_k = C_1\lambda_1^k + C_2\lambda_2^k, \qquad k \geq 1$$

其中 λ_1, λ_2 为 $\lambda^2 - (1+\theta_1^2)\lambda + \theta_1^2 = 0$ 的根，即 $\lambda_1 = 1, \lambda_2 = \theta_1^2$；$\Delta_k$ 的初始值可由直接计算得到：$\Delta_1 = 1+\theta_1^2$，$\Delta_2 = (1+\theta_1^2)^2 - \theta_1^2 = 1+\theta_1^2+\theta_1^4$，因此可以定出系数 $C_1 = (1-\theta_1^2)^{-1}$，$C_2 = -\theta_1^2(1-\theta_1^2)^{-1}$，于是

$$\Delta_k = (1-\theta_1^{2(k+1)})(1-\theta_1^2)^{-1}$$

最后解出 $\varphi_{11} = -\dfrac{\theta_1}{1+\theta_1^2}$ 和

$$\varphi_{kk} = \frac{(-1)^{k+1}(-\theta_1)^k}{\Delta_k}$$

$$= \frac{-(1-\theta_1^2)\theta_1^k}{1-\theta_1^{2k+2}}, \qquad k \geqslant 2$$

由于 MA(0, 1) 的可逆域为 $|\theta_1| < 1$，所以 $|\rho_1| = \left|\frac{\theta_1}{1+\theta_1^2}\right| < \frac{1}{2}$，这就是 MA(0, 1) 的允许域.

例 3. MA(0, 2) 序列 $w_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$

$$\gamma_0 = (1+\theta_1^2+\theta_2^2)\sigma_a^2, \quad \gamma_1 = (-\theta_1+\theta_1\theta_2)\sigma_a^2$$

$$\gamma_2 = -\theta_2\sigma_a^2, \quad \gamma_k = 0, \quad k \geqslant 3$$

$$\rho_1 = (-\theta_1+\theta_1\theta_2)(1+\theta_1^2+\theta_2^2)^{-1}$$

$$\rho_2 = -\theta_2(1+\theta_1^2+\theta_2^2)^{-1}, \quad \rho_k = 0, \quad k \geqslant 3$$

它的允许域为

$$\rho_2+\rho_1 > -\frac{1}{2}, \quad \rho_2-\rho_1 > -\frac{1}{2}, \quad \rho_1^2 < 4\rho_2(1-2\rho_2)$$

$$\left(\rho_2 \geqslant \frac{1}{6} \text{时}\right)$$

其具体推导方法与 § 6 例 4 的技巧类似，这种烦琐的初等论证我们不一一列举.

例 4. AR(1, 0) 序列 $w_t - \varphi_1 w_{t-1} = a_t$.

$$\gamma_k = \gamma_0\varphi_1^k; \quad \rho_k = \varphi_1^k, \quad k \geqslant 0$$

$$\varphi_{11} = \rho_1 = \varphi_1, \quad \varphi_{kk} = 0, \quad k \geqslant 2$$

它的允许域为 $|\rho_1| = |\varphi_1| < 1$.

例 5. AR(2, 0) 序列 $w_t - \varphi_1 w_{t-1} - \varphi_2 w_{t-2} = a_t$.

它的自相关函数已在前边例 1 中讨论过了. 它的偏相关函数十分简单,即

$$\varphi_{11} = \rho_1; \quad \varphi_{22} = \varphi_2, \quad \varphi_{21} = \varphi_1$$

$$\varphi_{kk} = 0, \quad \varphi_{k1} = \varphi_1, \quad \varphi_{k2} = \varphi_2, \quad \varphi_{kj} = 0, \quad k \geqslant j \geqslant 3$$

它的允许域为

$$-1 < \rho_1 < 1, \quad -1 < \rho_2 < 1, \quad \rho_1^2 < \frac{1}{2}(1+\rho_2)$$

例 6. ARMA(1, 1) 序列，$w_t - \varphi_1 w_{t-1} = a_t - \theta_1 a_{t-1}$.

利用求模型 (2.1.16 e)的自协方差函数的同样方法,可以得到

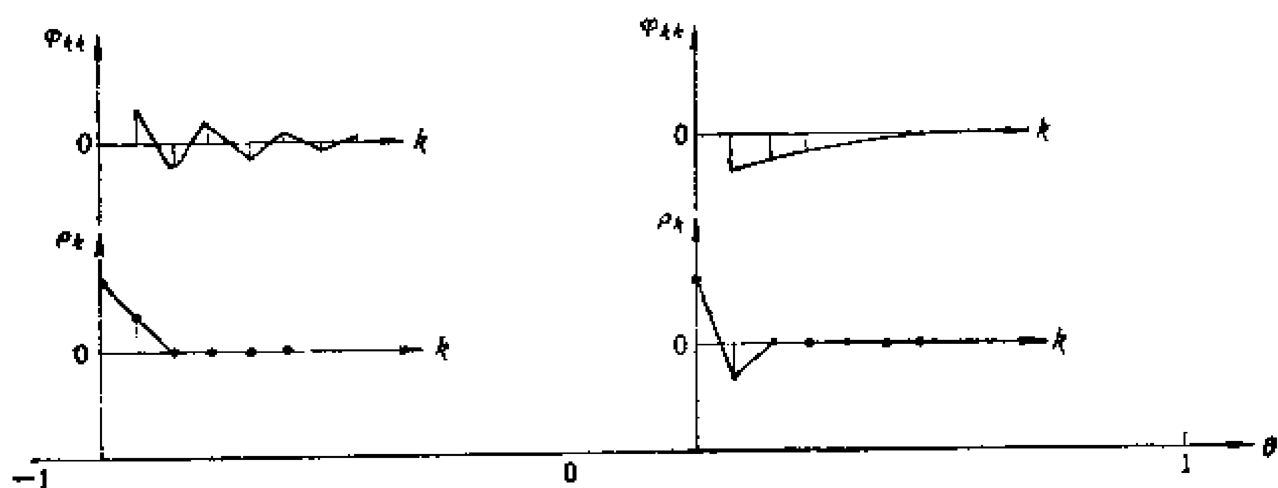


图 2.2.2(a)　MA(0,1) 自相关、偏相关函数

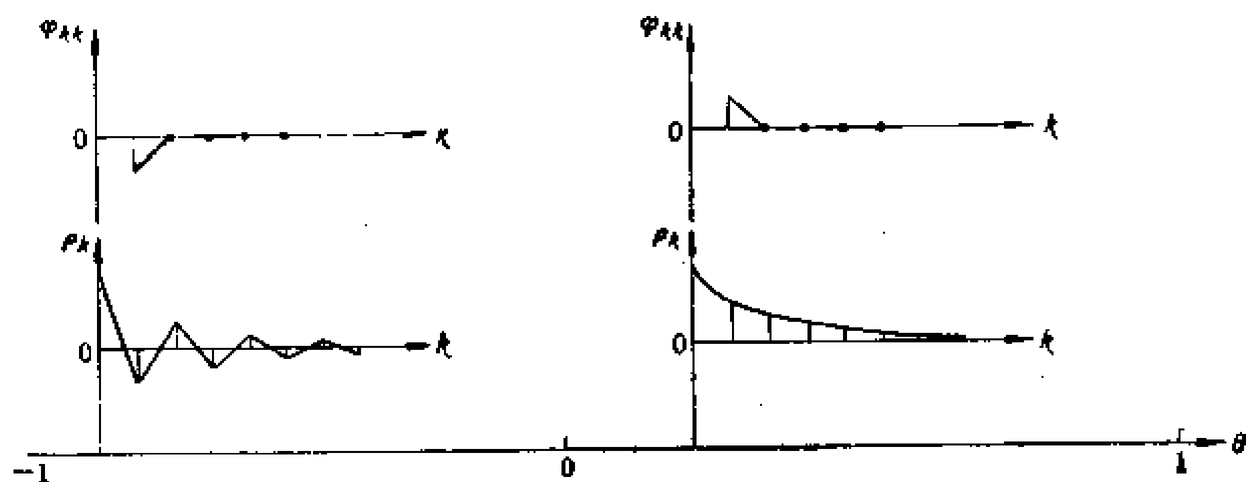


图 2.2.2(b)　AR(1,0) 的自相关、偏相关函数

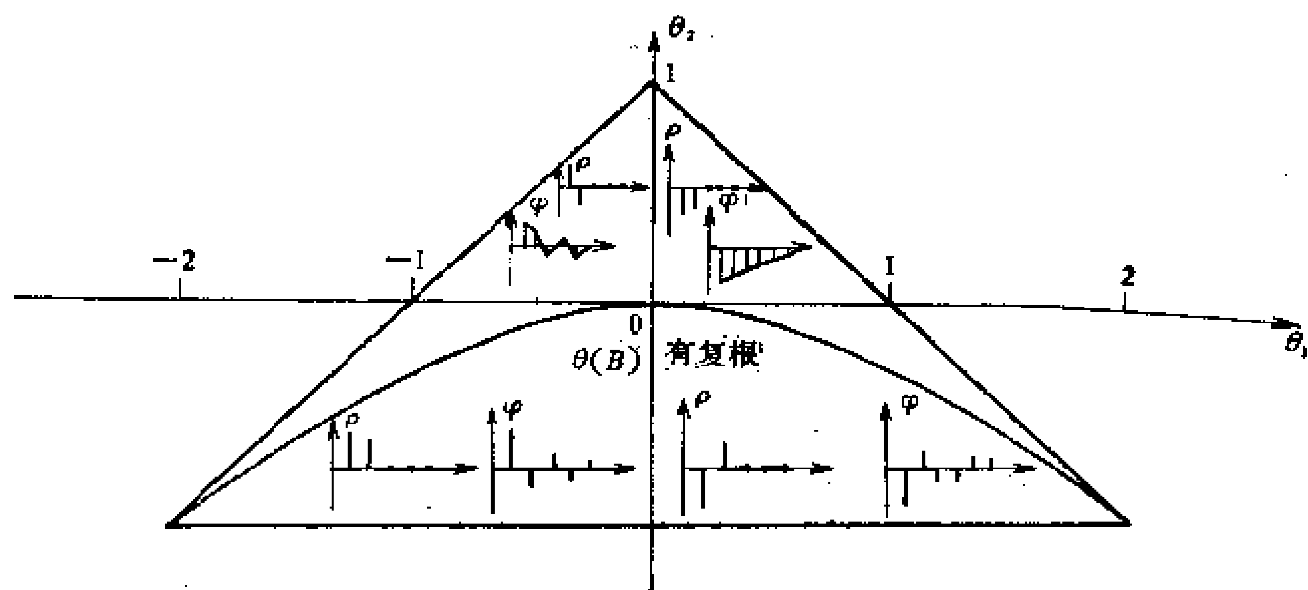


图 2.2.2(c)　MA(0,2) 的自相关、偏相关函数

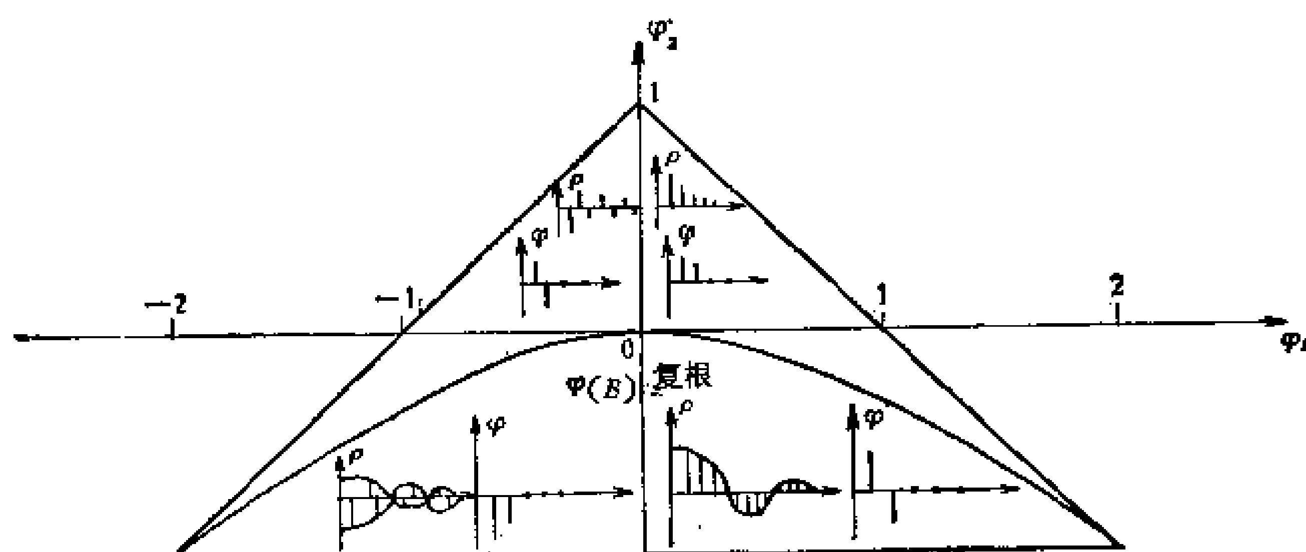


图 2.2.2(d)　AR(2,0) 的自相关、偏相关函数

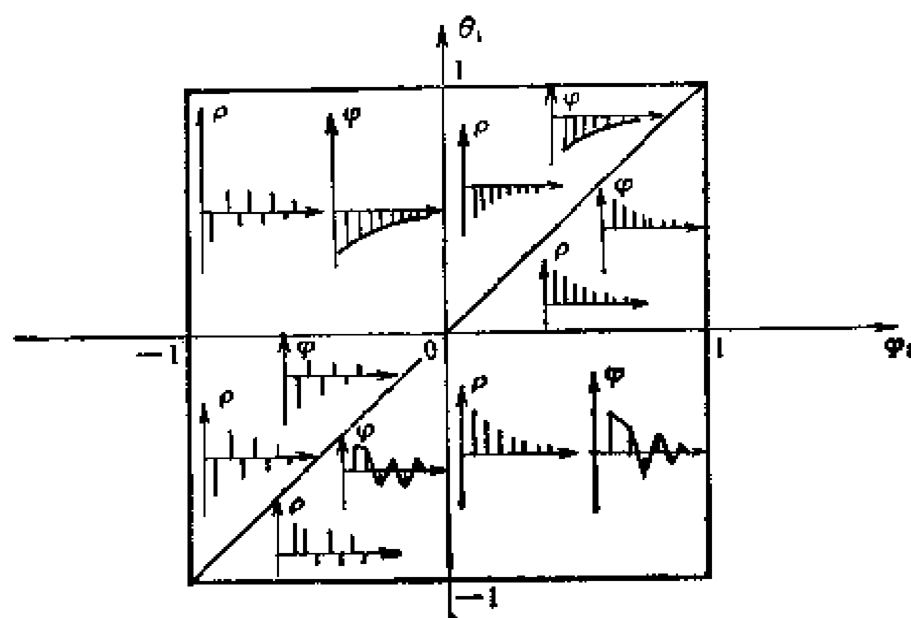


图 2.2.2(e)　ARMA(1,1) 自相关、偏相关函数

$$
\gamma_k = \begin{cases} \dfrac{1+\theta_1^2-2\varphi_1\theta_1}{1-\varphi_1^2}\sigma_a^2, & k=0 \\ \dfrac{(1-\varphi_1\theta_1)(\varphi_1-\theta_1)}{1-\varphi_1^2}\sigma_a^2\varphi_1^{k-1}, & k\geqslant 1 \end{cases}
$$

$$
\rho_k = \begin{cases} 1, & k=0 \\ \dfrac{(1-\varphi_1\theta_1)(\varphi_1-\theta_1)}{1+\theta_1^2-2\varphi_1\theta_1}\varphi_1^{k-1}, & k\geqslant 1 \end{cases}
$$

它的允许域为

$$|\rho_2| < |\rho_1|, \quad \rho_2 > \rho_1(2\rho_1 + 1) \qquad (\rho_1 < 0 \text{时})$$

$$\rho_2 > \rho_1(2\rho_1 - 1) \ (\rho_1 > 0 \text{时})$$

当 φ, θ 取不同值时，ρ_k, φ_{kk} 的各种形状，可参见图 2.2.2.

到此为止，我们已经介绍了 ARMA 模型的分类及其性质，ARMA 序列的自相关与偏相关函数的各种性质. 现在我们把这些内容的主要部分用以下的一览表综合起来，这样比较醒目，而且便于供下一章使用.

表 2.2.1　ARMA 序列的分类性质一览表

类别 \ 模型	AR $(p, 0)$	MA $(0, q)$	ARMA (p, q)
模型方程	$\varphi(B)w_t = a_t$	$w_t = \theta(B)a_t$	$\varphi(B)w_t = \theta(B)a_t$
平稳条件	$\varphi(\omega) = 0$ 的根之模 >1	无条件	$\varphi(\omega) = 0$ 的根之模 >1
可逆条件	无条件	$\theta(\omega) = 0$ 的根之模 >1	$\theta(\omega) = 0$ 的根之模 >1
传递形式	$w_t = \varphi^{-1}(B)a_t$	$w_t = \theta(B)a_t$	$w_t = \varphi^{-1}(B)\theta(B)a_t$
逆转形式	$a_t = \varphi(B)w_t$	$a_t = \theta^{-1}(B)w_t$	$a_t = \theta^{-1}(B)\varphi(B)w_t$
ψ_k 序列	拖尾	截尾	拖尾
π_k 序列	截尾	拖尾	拖尾
自相关函数	拖尾	截尾	拖尾
偏相关函数	截尾	拖尾	拖尾

§3. 一类非平稳序列—— ARIMA 序列

在许多实际问题中，所考虑的随机序列 z_t 有时并不近似为平稳序列. 比如，在绪论中所举的用电负荷的例子，当我们长期考查这个序列时，就会发现它并不总是稳定在同一水平上，因为我们的工业在发展，用电量和供电能力随之也在上升. 再看第一章所举的第六个伪随机序列的数字例子（由模型 (2.1.16 f) 造出），它很明显地与其它五个平稳序列不同. 但是，如果将这个数列进行一次差分，得到 ∇z_t（我们将 z_t 和 ∇z_t 一起画在图 2.3.1 里），这时，我们会发现 ∇z_t 几乎稳定在零附近，即可以近似看做平稳序列. 这是为什么呢？从 (2.1.16 f) 式可以得到答案. 在第二章 § 1 中，我们

曾说过 z_t 是按(2.1.16 f)式构造的，并且随后又指出 z_t 并非 AR(2, 0)序列. 现在改写(2.1.16 f)式为

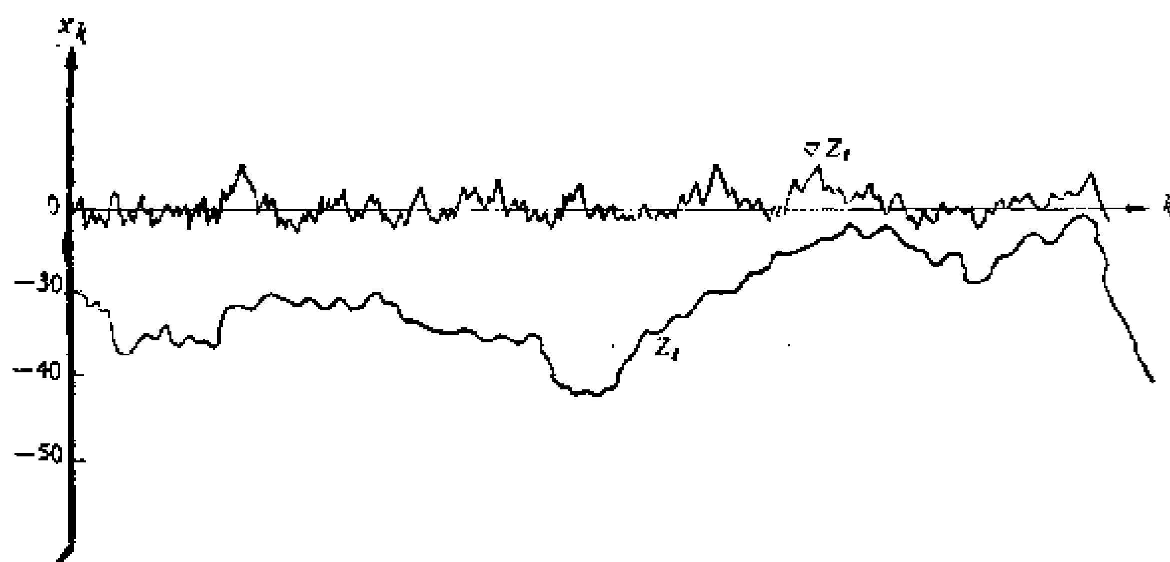


图 2.3.1 (2.1.16f)式的模型数据列与差分值

$$(z_k - z_{k-1}) - 0.5(z_{k-1} - z_{k-2}) = a_k \qquad (2.3.1)$$

若记 $\nabla z_k = z_k - z_{k-1} = w_k$，则上式又可写成

$$w_k - 0.5w_{k-1} = a_k \qquad (2.3.2)$$

这恰好等于模型(2.1.16 c)式. 这意味着，z_t 的差分序列 $\nabla z_t = w_t$ 是 AR(1, 0) 序列. 当然，由此不难想象，对某些非平稳序列 z_t，∇z_t 也还不是平稳的，但 $\nabla^2 z_t$ 是平稳的，等等. 比如绪论中例 2 的 x_k，它的三次差分才是平稳序列.

我们把上述处理方法推广到一般情形. 假定 z_t 为一非平稳序列，但是存在一正整数 d，使得

$$\nabla^d z_t = w_t \qquad t > d \qquad (2.3.3)$$

而 w_t 是 ARMA(p, q)序列，其中 ∇^d 是 d 阶差分算子

$$\nabla^d = (1 - B)^d = 1 - C_1^d B + C_2^d B^2 + \cdots + (-1)^{d-1} C_{d-1}^d B^{d-1} + (-1)^d B^d \qquad (2.3.4)$$

式中 $C_r^d = \dfrac{d!}{r!(d-r)!}$. 由于我们只考虑 z_t 的方差 Ez_t^2 为有穷值的情况，所以，对这类非平稳序列，我们假定 t 从 1 开始 z_t 才有

定义，并且假定 z_t 的前 d 个随机变量 $z_1, z_2, \cdots, z_d$ 是均值为零，方差有穷，且与 w_t 不相关，因而与 a_t 也不相关．这样的 z_t 序列我们称之为 ARMA 序列的 d 阶求和序列，并用记号 ARIMA(p, d, q) 表示．其中 d 为求和阶数，p, q 分别为 $\nabla^d z_t = w_t$ 的自回归和滑动平均阶数．

这样的非平稳序列可以用它的初值 $z_1, z_2, \cdots, z_d$ 及平稳序列 w_t 表达．事实上，因为差分运算的逆运算就是求和运算，所以由 (2.3.3) 式和 z_t 的初值可以得到

$$\begin{aligned} z_t &= \sum_{i=0}^{d-1} C_i^{t-d+i-1} \nabla^i z_d + \sum_{j=1}^{t-d} C_{d-1}^{t-j-1} w_{d+j} \\ &= \sum_{i=0}^{d-1} C_i^{t-k+i-1} \nabla^i z_k + \sum_{j=1}^{t-k} C_{d-1}^{t-k-j+d-1} w_{k+j}, \quad t > k \geqslant d \end{aligned} \tag{2.3.5}$$

我们不去仔细推导这些表达式，仅考虑两种简单而常用的特殊情况．

$d = 1$ 时，可表

$$\begin{aligned} z_t &= z_t - z_{t-1} + z_{t-1} - \cdots + z_2 - z_1 + z_1 \\ &= z_1 + \sum_{j=1}^{t-1} (z_{j+1} - z_j) = z_1 + \sum_{j=1}^{t-1} w_{1+j} \\ &= z_k + \sum_{j=1}^{t-k} w_{k+j}, \qquad t > k \geqslant d \end{aligned} \tag{2.3.6}$$

$d = 2$ 时，同上有

$$z_t = z_1 + \sum_{j=1}^{t-1} \nabla z_{j+1}$$

同理又有

$$\nabla z_{j+1} = \nabla z_2 + \sum_{i=1}^{j-1} \nabla^2 z_{i+2}$$

以此代入前式可得

$$z_t = z_2 + (t-2)(z_2 - z_1) + \sum_{j=1}^{t-2} (t-j-1) w_{2+j}$$

$$= z_k + (t-k)(z_k - z_{k-1}) + \sum_{j=1}^{t-k} (t-k-j+1) w_{k+j}$$

$$t > k \geqslant 2 \qquad (2.3.7)$$

对于非平稳 ARIMA(p, d, q) 序列，它的识别估计和预报控制等问题，可以通过对它进行 d 阶差分化成平稳 ARMA(p, q) 序列的相应问题来解决. 可见这是将平稳时间序列分析方法向非平稳序列推进了一步，但并非本质性的推进.

最后，还须指出，有时我们也将(2.3.3)式形式地写成

$$\varphi(B)\nabla^d z_t = \theta(B) a_t \qquad t > d \qquad (2.3.8)$$

但是严格地说，还是用(2.3.5)式(或(2.3.6)，(2.3.7)式)更为恰当. 那样 z_t 可看成是由 $z_1, z_2, \cdots, z_d$ 和 w_t 按(2.3.5)式产生的非平稳序列，其中 w_t 是如下的 ARMA(p, q) 模型的平稳解

$$\varphi(B) w_t = \theta(B) a_t, \quad t = \cdots, -1, 0, 1, \cdots$$

而且与 $z_1, z_2, \cdots, z_d$ 不相关. 对求和序列 z_t 的这种理解，在今后各章中将多次用到.

第三章　模型的初步识别与参数的矩估计

上一章，我们讨论了 ARMA 序列的自相关与偏相关函数的性质，及其与模型参数之间的关系．在第二章 §1 又曾指出实际中的大量随机序列，并不能用直接方法（或称物理方法）建立它们的模型，从物理意义上看甚至于它们本身并不是线性模型．这时，所谓根据它们的样本数据建立线性模型，不如说是寻找一种与它们尽可能“等价”的 ARMA 序列，作为实际序列的近似模型．前面也已说过，我们之所以限于处理这类模型，一方面是由于 ARMA 序列的谱密度是有理的，它们能逼近任何连续谱密度；另一方面是由于在实际应用中，这类模型有很多方便之处．以下三章的内容都是围绕着这样的问题展开的：如果 w_t 是某一未知其模型的 ARMA 序列，现在获得了它的一段样本数据 $w_1, w_2, \cdots, w_N$，如何根据这 N 个数值对 w_t 的模型做出估计，或者说得更具体些，对模型的阶数 p, q 和参数 $\boldsymbol{\varphi}, \boldsymbol{\theta}$ 与 σ_a^2 做出判断与估计．一般，我们把对 p, q 的判断称为模型识别，也就是初步确定序列模型的形状；在 p, q 被判定后，估计相应的 $\boldsymbol{\varphi}, \boldsymbol{\theta}$ 和 σ_a^2，称为参数估计．解决这些问题的方法属于时间序列的内容．一旦作出了序列模型的识别和估计，序列的自相关、偏相关函数及谱密度的估计就可以由这个模型算出．对于很多实际问题里的随机序列，我们总假定，必要时经过某种处理（例如差分或“季节型”差分，详见第六章）后，可以看作 ARMA 序列，然后按上述方法分析，并以估计所得的模型做为真实序列的近似描述．

解决上述识别与估计问题，主要依据前一章的理论分析结果．本章将介绍一种初步识别方法和参数的矩估计，这些方法简单易懂，但是其精度稍差．对于某些实际情况，如果计算条件允许，并且希望尽量提高精度时，可根据第四章的方法求参数的精估计，再

根据第五章的内容，检验和改进初步识别的结果. 不过这两章的方法较烦，计算量也大得多.

§1. 样本自相关与样本偏相关函数

假定得到了数据 $z_1, z_2, \cdots, z_N$，它们是随机序列 z_t 的一段样本值，N 称为样本长度. 为要根据这段样本数据估计 z_t 的模型，先要设法消去 z_t 的均值项，至于去掉均值的方法，将在以后提到. 在本节中，为了统一符号，暂时假定 $Ez_t = 0$. 这时定义 z_t 的样本自协方差 $\hat{\gamma}_k$ 与样本自相关函数 $\hat{\rho}_k$ 分别为：

$$\hat{\gamma}_k = \hat{\gamma}_{-k} = \frac{1}{N}\sum_{l=1}^{N-k} z_l z_{l+k} \quad k = 0, 1, 2, \cdots, (N-1) \tag{3.1.1}$$

$$\hat{\rho}_k = \hat{\rho}_{-k} = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} \qquad k = 0, 1, 2, \cdots, (N-1) \tag{3.1.2}$$

在其它地方，也有采用如下定义，

$$\hat{\gamma}_k^* = \hat{\gamma}_{-k}^* = \frac{1}{N-k}\sum_{l=1}^{N-k} z_l z_{l+k} \quad k = 0, 1, 2, \cdots, (N-1) \tag{3.1.3}$$

$$\hat{\rho}_k^* = \hat{\rho}_{-k}^* = \frac{\hat{\gamma}_k^*}{\hat{\gamma}_0^*} \qquad k = 0, 1, 2, \cdots, (N-1) \tag{3.1.4}$$

可见 $\hat{\gamma}_k = \dfrac{N-k}{N}\hat{\gamma}_k^*$，因此，$\hat{\gamma}_k$ 与 $\hat{\gamma}_k^*$ 是渐近相等的. 但是，$\hat{\gamma}_k$ 是非负定序列，$\hat{\gamma}_k^*$ 则不一定如此. 为了证明这一点，对于 $l > N$ 或 $l \leqslant 0$ 暂记 $z_l = 0$，于是对于任意 m 个实数 $g_1, g_2, \cdots, g_m$，

$$\begin{aligned}\sum_{i,j=1}^{m} g_i g_j \hat{\gamma}_{j-i} &= \frac{1}{N}\sum_{i,j=1}^{m} g_i g_j \sum_{l=1}^{N-|j-i|} z_l z_{l+|j-i|} \\ &= \frac{1}{N}\sum_{i,j=1}^{m} g_i g_j \sum_{l=-\infty}^{\infty} z_l z_{l+|j-i|} = \frac{1}{N}\sum_{i,j=1}^{m} g_i g_j \sum_{l=-\infty}^{\infty} z_l z_{l+j-i} \\ &= \frac{1}{N}\sum_{i,j=1}^{m} g_i g_j \sum_{l=-\infty}^{\infty} z_{l+i} z_{l+j} = \frac{1}{N}\sum_{l=-\infty}^{\infty}\left(\sum_{i=1}^{m} g_i z_{l+i}\right)^2 \geqslant 0\end{aligned}$$

这就证明了 $\hat{\gamma}_k$ 序列的非负定性.

如果 z_t 是 ARMA 序列，(注意若不加特别说明，恒假定随机序列是正态的)，那么 $\hat{\gamma}_k$ 和 $\hat{\rho}_k$ 做为 γ_k 和 ρ_k 的估计量，具有以下性质：

1. 它们是相容估计和渐近无偏估计，即

$$\lim_{N\to\infty}\hat{\gamma}_k \overset{p}{=} \gamma_k, \qquad \lim_{N\to\infty}\hat{\rho}_k \overset{p}{=} \rho_k \tag{3.1.5}$$

$$\lim_{N\to\infty}E\hat{\gamma}_k = \gamma_k, \qquad \lim_{N\to\infty}E\hat{\rho}_k = \rho_k \tag{3.1.6}$$

(3.1.5) 式中"$\overset{p}{=}$"表示依概率相等.

2. 它们具有渐近正态性，即存在两个趋于 ∞ 的数列 $b_N^{(1)}$, $b_N^{(2)}$，分别使得

$$b_N^{(1)}(\hat{\gamma}_k - \gamma_k) \sim \mathcal{N}(0,1), \quad b_N^{(2)}(\hat{\rho}_k - \rho_k) \sim \mathcal{N}(0,1)$$

3. 它们的误差方差分别渐近为

$$\begin{aligned}\operatorname{var}(\hat{\gamma}_k) &= E\tilde{\gamma}_k^2 = E(\hat{\gamma}_k - \gamma_k)^2 \\ &\approx \frac{1}{N}\sum_{l=-\infty}^{\infty}(\gamma_l^2 + \gamma_{l+k}\gamma_{l-k})\end{aligned} \tag{3.1.7}$$

$$\begin{aligned}\operatorname{var}(\hat{\rho}_k) &= E\tilde{\rho}_k^2 = E(\hat{\rho}_k - \rho_k)^2 \\ &\approx \frac{1}{N}\sum_{l=-\infty}^{\infty}(\rho_l^2 + \rho_{l+k}\rho_{l-k} + 2\rho_k^2\rho_l^2 - 4\rho_k\rho_l\rho_{l-k})\end{aligned} \tag{3.1.8}$$

(3.1.8) 式称为 Bartlett 公式，这些公式的证明可见本书附录 § 2.
特别由 (3.1.7) 式可知，当 z_t 为白噪声时，

$$\operatorname{var}(\hat{\gamma}_0) \approx \frac{2\gamma_0^2}{N}, \quad \operatorname{var}(\hat{\gamma}_k) \approx \frac{\gamma_0^2}{N}, \quad k > 0 \tag{3.1.9}$$

当 z_t 为 MA(0, q) 序列时，

$$\operatorname{var}(\hat{\gamma}_0) \approx \frac{2\gamma_0^2}{N}\left(1 + 2\sum_{l=1}^{q}\rho_l^2\right)$$

$$\operatorname{var}(\hat{\gamma}_k) \approx \frac{\gamma_0^2}{N}\left(1 + 2\sum_{l=1}^{q}\rho_l^2\right), \quad k > q \tag{3.1.10}$$

当 z_t 为 AR(1, 0) 序列时，$\gamma_l = \gamma_0\varphi_1^{|l|}$，以此代人 (3.1.10) 式可算出

$$
\begin{cases}
\operatorname{var}(\hat{\gamma}_0) \approx \dfrac{2\gamma_0^2}{N} \dfrac{1+\varphi_1^2}{1-\varphi_1^2} \\
\operatorname{var}(\hat{\gamma}_k) \approx \dfrac{\gamma_0^2}{N}\left(\dfrac{(1+\varphi_1^2)(1+\varphi_1^{2k})}{1-\varphi_1^2} + 2k\varphi_1^{2k}\right), k > 0
\end{cases}
\tag{3.1.11}
$$

以上诸式告诉我们，$\hat{\gamma}_k$ 和 $\hat{\rho}_k$ 的渐近误差方差都是与 $\frac{1}{N}$ 成正比例的，而其比例常数又和真值 γ_l 或 ρ_l 有关．显然当 $N \to \infty$ 时，诸误差方差都趋于零，由此可推出 (3.1.5) 式和 (3.1.6) 式．对于 MA(0, 1) 序列，因为 $|\rho_1| < \frac{1}{2}$，$\rho_k = 0$，$k > 1$，所以 $\operatorname{var}(\hat{\gamma}_0)$ 和 $\operatorname{var}(\hat{\gamma}_k)$ ($k > 1$) 的主项分别被 $\frac{3\gamma_0^2}{N}$ 和 $\frac{1.5\gamma_0^2}{N}$ 所控制，即它们趋于零的速度不会比 $\frac{3\gamma_0^2}{N}$ 和 $\frac{1.5\gamma_0^2}{N}$ 更慢．对于 AR(1, 0) 序列，由 (3.1.11) 式可以看出，$\operatorname{var}(\hat{\rho}_k)$ 显著地依赖于参数 φ_1，φ_1 越接近它的平稳域边界 (± 1)，$\operatorname{var}(\hat{\gamma}_k)$ 趋于零的速度越慢．例如当 $\varphi_1 = 0.1$ 时，$\operatorname{var}\hat{\gamma}_0 \approx \frac{2.03\gamma_0^2}{N}$；而当 $\varphi_1 = 0.9$ 时，$\operatorname{var}(\hat{\gamma}_k) = \frac{19.05\gamma_0^2}{N}$，不仅如此，(3.1.11)式还找不到与 φ_1 无关的上界．对于 ARMA 序列来说也有类似的情况．

在实际应用中，我们希望通过增加样本长度 N 以提高估计精度．为了保证上述各估计量的均方误差达到一定的精度要求（比如要求 $\operatorname{var}(\hat{\gamma}_0) < 0.01$，或者要求 $\operatorname{var}(\hat{\gamma}_0) < 0.01\gamma_0^2$；前者叫绝对精度，后者叫相对精度），可以用 (3.1.7) 式和 (3.1.8) 式计算 N 需要多大．但是，这些公式中的 ρ_k 或 γ_k 仍是未知的，所以并不能准确定出多大的 N 才能满足预定的精度．为此，可采用以下不断实践不断改进的方法进行摸索．假定开始获得了 N_1 个数据，由 (3.1.1)，(3.1.2) 式算出适当长度的 $\hat{\gamma}_k$，$\hat{\rho}_k$，直到它们接近于零．为了对 $\hat{\gamma}_k$，$\hat{\rho}_k$ 的方差做出估计，利用 (3.1.7)，(3.1.8) 式，在这两式的右方将 γ_k 和 ρ_k 换成 $\hat{\gamma}_k$ 和 $\hat{\rho}_k$，并且将无穷和换成有穷和，由于 γ_k，ρ_k 依指数律收敛于零，这样的近似是可行的．如果所算出的 $\operatorname{var}(\hat{\gamma}_k)$ 或 $\operatorname{var}(\hat{\rho}_k)$ 达到了预定的精度要求，就认为取样长

度 N_1 是合适的，否则应设法增加取样数据．比如根据现有的 $\mathrm{var}(\hat{\gamma}_k)$ 或 $\mathrm{var}(\hat{\rho}_k)$ 的大小和精度要求，试着增加到 N_2，然后再重复作出新的估计，计算出新的 $\mathrm{var}(\hat{\gamma}_k)$ 和 $\mathrm{var}(\hat{\rho}_k)$ 之值，并检查是否满足预定的精度．按这个程序做下去，直到合适的 N 为止．通常只要重复一两次便能确定出 N 来，如果总定不下来，则序列 z_t 可能不是 ARMA 序列．

有了样本自相关函数，便可以按照 (2.2.21) 式定义样本偏相关函数了．解方程

$$\begin{pmatrix} 1 & \hat{\rho}_1 & \cdots & \hat{\rho}_{k-1} \\ \hat{\rho}_1 & 1 & \cdots & \hat{\rho}_{k-2} \\ \cdots & \cdots & \cdots & \cdots \\ \hat{\rho}_{k-1} & \hat{\rho}_{k-2} & \cdots & 1 \end{pmatrix} \begin{pmatrix} \hat{\varphi}_{k1} \\ \hat{\varphi}_{k2} \\ \vdots \\ \hat{\varphi}_{kk} \end{pmatrix} = \begin{pmatrix} \hat{\rho}_1 \\ \hat{\rho}_2 \\ \vdots \\ \hat{\rho}_k \end{pmatrix} \tag{3.1.12}$$

得到 $\hat{\varphi}_{kk}(k=1,2,\cdots)$，这就是 z_t 的样本偏相关函数．注意，$\hat{\varphi}_{kk}$ 也可以按 (2.2.23) 式递推计算，只须将那里的 ρ_j 换成 $\hat{\rho}_j$．当 z_t 为 AR$(p,0)$ 序列时，根据附录 §5 的结果，$\hat{\varphi}_{kk}$ 具有以下性质：

1. $\hat{\varphi}_{kk}$ 是 φ_{kk} 的渐近无偏估计和相容估计，即

$$\lim_{N\to\infty} E\hat{\varphi}_{kk} = \varphi_{kk}, \qquad \lim_{N\to\infty} \hat{\varphi}_{kk} \overset{p}{=} \varphi_{kk}$$

特别

$$\lim_{N\to\infty} E\hat{\varphi}_{kk} = \begin{cases} \varphi_k, & 当\ k=p \\ 0, & 当\ k>p \end{cases}$$

2. $\hat{\varphi}_{kk}$ 具有渐近正态性质．即存在趋于 ∞ 的数列 $b_N^{(3)}$ 使得

$$b_N^{(3)}\tilde{\varphi}_{kk} = b_N^{(3)}(\hat{\varphi}_{kk} - \varphi_{kk}) \sim \mathcal{N}(0,1)$$

3. 当 $k, k' > p, k \neq k'$ 时

$$E\tilde{\varphi}_{kk}^2 = E\tilde{\varphi}_{k'k'}^2 \approx \frac{1}{N}, \quad NE\tilde{\varphi}_{kk}\tilde{\varphi}_{k'k'} \approx 0 \tag{3.1.13}$$

对于一般的 ARMA(p,q) 序列，$\hat{\varphi}_{kk}$ 也具有与此相类似的性质，不过主要用到的是 AR$(p,0)$ 序列的偏相关函数．

§2. 模型的初步识别方法

本节讨论如何利用前一节的样本自相关与样本偏相关函数，

来判断 z_t 的模型阶数，即 p 和 q 之值；若 z_t 是求和模型，还要判断 d 的值．初步识别方法的要领是：

1. 识别的依据：根据第二章中 ARMA 序列自相关与偏相关函数的性质（见第二章 § 2 表 2.2.1），若其样本自相关函数 $\hat{\rho}_k$ 在 $k > q$ 后截尾，很自然地判断 z_t 是 MA(0, q) 序列．类似地，若 $\hat{\varphi}_{kk}$ 在 $k > p$ 以后截尾，则判断 z_t 是 AR(p, 0) 序列．若 $\hat{\rho}_k$，$\hat{\varphi}_{kk}$ 都不截尾，又被负指数型的数列所控制（即拖尾的），则应判断其为 ARMA 序列，但尚不能定出阶数．

2. $\hat{\rho}_k$ 或 $\hat{\varphi}_{kk}$ 截尾性的判断：理论自相关和偏相关函数 ρ_k 和 φ_{kk} 的截尾性，是指它们从某个值 q 或 p 以后全为零．由于 $\hat{\rho}_k$，$\hat{\varphi}_{kk}$ 是 ρ_k 和 φ_{kk} 的估计值，它们必然有误差，所以，即使 z_t 是 MA(0, q)序列，当 $k > q$ 后，$\hat{\rho}_k$ 也不会全为零，而只是在零上下起伏．根据上一节所述的 $\hat{\rho}_k$ 的渐近性质，若真值 ρ_k 在 $k > q$ 以后截尾，则 $\rho_k = 0$，而 $\hat{\rho}_k$ 的分布渐近为 $\mathscr{N}\left(0, \frac{1}{N}\left(1 + 2\sum_{l=1}^{q} \hat{\rho}_l^2\right)\right)$．依渐近正态分布律可知

$$P\left(|\hat{\rho}_k| \leqslant \frac{1}{\sqrt{N}}\left(1 + 2\sum_{l=1}^{q} \hat{\rho}_l^2\right)^{\frac{1}{2}}\right) = 68.3\%$$

$$P\left(|\hat{\rho}_k| \leqslant \frac{2}{\sqrt{N}}\left(1 + 2\sum_{l=1}^{q} \hat{\rho}_l^2\right)^{\frac{1}{2}}\right) = 95.5\%$$

对每个 $q > 0$ 都可以检验 $\hat{\rho}_{q+1}, \hat{\rho}_{q+2}, \cdots, \hat{\rho}_{q+M}$（$M$ 在经验上一般取 $\sqrt{N}$ 或 $\frac{N}{10}$ 左右）中满足 $|\hat{\rho}_k| \leqslant \frac{1}{\sqrt{N}}\left(1 + 2\sum_{l=1}^{q} \hat{\rho}_l^2\right)^{\frac{1}{2}}$ 或 $|\hat{\rho}_k| \leqslant \frac{2}{\sqrt{N}}\left(1 + 2\sum_{l=1}^{q} \hat{\rho}_l^2\right)^{\frac{1}{2}}$ 的比例是否达到 68.3% 或 95.5%，若 $q = 1, 2, \cdots, q' - 1$ 都没有达到，而 $q = q'$ 达到了，我们就说 $\hat{\rho}_k$ 在 q' 以后截尾，因而按照要领 1 判断 z_t 为 MA 序列，并且 q' 就称为 q 的初步识别值．为了简化符号，一般不用 q' 而仍用 q 来表示．由于 $\hat{\rho}_k(k > q)$ 对于不同的 k 尚有某些相关，仅当不同的 k 之差超过 q 后才不相关（见附录 § 2），所以这种检验只能做

为一种近似手段．对于 $\hat{\varphi}_{kk}$ 也有与上类似的办法．当然，这样的判断方法有时会把 MA 序列判断成其它类型，有时也会把 AR 序列判断成其它类型，而且即使类型判断无误，其阶数也可能有误．这属于判断精度问题，将在本章 § 4 中专门讨论．

3. 求和阶数 d 的识别：若 $\hat{\rho}_k$ 和 $\hat{\varphi}_{kk}$ 依要领 2 的分析不但都不截尾，而且（至少有一个）下降趋势很慢，则可认为它们不是拖尾的，即不能被负指数列控制．这时，可重新计算并按要领 1 和 2 分析 $\nabla z_k(k=2, 3, \cdots, N)$ 的样本自相关与样本偏相关函数．若这两个函数仍都不截尾，而且（至少有一个）下降很慢，则可再考虑对 $\nabla^2 z_k(k=3, 4, \cdots, N)$ 进行分析，直到某一次 $\nabla^d z_k(k=d+1, \cdots, N)$ 的样本自相关或样本偏相关函数为截尾或都为拖尾为止．这时 d 即初步判断为 z_t 的求和阶数．当然，d 与真值可能有出入．在某些实际应用中，若能根据数据的来源直接提供 d 的值，则不必对它进行估计．

4. 混合模型的定阶：若经要领 1，2，3 已判断 $w_t=\nabla^d z_t$ 属于 ARMA 型，由要领 1 知道，这时还不能定出 p 和 q 的值，只是知道它们都不为零．对于这种混合模型，识别 p, q 的办法可以从低阶到高阶逐个取 (p, q) 为 $(1, 1)$，$(1, 2)(2, 1)$ ……等值尝试．所谓尝试，就是先认定 (p, q) 为某值（如 $(1,1)$），然后进行下一步的参数估计，并定出估计模型来，再经过第五章的方法检验这个估计模型是否被接受，也就是与原序列符合得好不好；若不被接受，就调整 (p, q) 的尝试值，再重新做参数估计和检验，直到被接受为止．初看起来这个方法过于烦琐，其实，即使是判断为 AR 或 MA 序列时，也要通过检验手续后才能放心使用．另外，对于混合模型而言，实际应用中的阶数 (p, q) 一般都比较低．除此以外，也有人主张尽量用 AR 模型来逼近真实序列，而少采用 MA 或 ARMA 模型．还有一些其它定阶方法将在以后的章节中提到．

5. 去掉 z_t 的均值项：若我们分析的时间序列 z_t 包含非随机的均值项 f_t，那末在计算 z_t 的样本自协方差函数、偏相关函数以及对它进行模型识别和参数估计时，一定先要设法将均值项去掉．

表 3.2.1 (2.1.16)诸

序列名称 \ $\hat{\rho}_k$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$	$\hat{\rho}_4$	$\hat{\rho}_5$	$\hat{\rho}_6$	$\hat{\rho}_7$	$\hat{\rho}_8$
(2.1.16 a)	−0.34	−0.05	0.09	−0.14	0.08	0.04	−0.06	0.04
(2.1.16 b)	−0.59	0.10	0.04	−0.07	0.07	−0.05	0.04	−0.05
(2.1.16 c)	0.56	0.30	0.17	0.05	0.07	0.05	−0.02	−0.05
(2.1.16 d)	0.80	0.59	0.42	0.32	0.25	0.17	0.10	0.05
(2.1.16 e)	−0.23	−0.13	−0.06	0.02	0.03	0.07	−0.11	0.07
(2.1.16 f)	0.99	0.98	0.96	0.94	0.92	0.90	0.89	0.86
(2.1.16 f) 差分	0.54	0.33	0.18	0.14	0.10	0.17	0.11	0.08

表 3.2.2 (2.1.16) 诸

序列名称 \ $\hat{\varphi}_{kk}$	$\hat{\varphi}_{11}$	$\hat{\varphi}_{22}$	$\hat{\varphi}_{33}$	$\hat{\varphi}_{44}$	$\hat{\varphi}_{55}$	$\hat{\varphi}_{66}$	$\hat{\varphi}_{77}$
(2.1.16 a)	−0.34	−0.19	0.01	−0.12	0.00	0.05	0.00
(2.1.16 b)	−0.59	−0.39	−0.20	−0.19	−0.10	−0.10	−0.03
(2.1.16 c)	0.56	−0.02	0.02	−0.07	0.10	−0.03	−0.07
(2.1.16 d)	0.80	−0.16	−0.00	0.08	−0.03	−0.06	−0.02
(2.1.16 e)	−0.23	−0.19	−0.15	−0.07	−0.02	0.07	−0.07
(2.1.16 f)	0.99	−0.17	−0.09	−0.09	−0.02	−0.01	−0.06
(2.1.16 f) 差分	0.54	0.05	−0.02	0.05	0.00	0.13	−0.04

序列的样本自相关函数

$\hat{\rho}_9$	$\hat{\rho}_{10}$	$\hat{\rho}_{11}$	$\hat{\rho}_{12}$	$\hat{\rho}_{13}$	$\hat{\rho}_{14}$	$\hat{\rho}_{15}$	$\hat{\rho}_{16}$	$\hat{\gamma}_0$	$\hat{\mu}$
−0.08	−0.02	0.08	−0.06	0.05	−0.07	0.04	0.05	1.23	0.001
0.10	−0.16	0.17	−0.10	0.00	0.05	−0.06	−0.02	2.25	−0.002
−0.09	−0.05	0.02	0.01	0.02	0.00	0.05	0.09	1.57	0.001
0.03	0.03	0.03	0.00	−0.05	−0.07	−0.08	−0.04	2.72	0.136
0.04	−0.12	0.06	0.04	−0.03	0.01	−0.02	0.09	1.19	−0.003
0.84	0.82	0.80	0.77	0.75	0.73	0.71	0.68	237.00	−27.44
0.03	0.04	−0.02	−0.05	−0.07	−0.06	−0.04	−0.07	1.34	0.140

序列的样本偏相关函数

$\hat{\varphi}_{88}$	$\hat{\varphi}_{99}$	$\hat{\varphi}_{10,10}$	$\hat{\varphi}_{11,11}$	$\hat{\varphi}_{12,12}$	$\hat{\varphi}_{13,13}$	$\hat{\varphi}_{14,14}$	$\hat{\varphi}_{15,15}$	$\hat{\varphi}_{16,1}$
0.01	−0.08	−0.07	0.01	−0.02	0.03	−0.07	0.02	0.05
−0.08	0.07	−0.10	0.04	0.01	−0.03	−0.00	−0.00	−0.12
−0.02	−0.05	−0.05	0.05	−0.03	0.02	−0.02	0.01	0.02
0.02	0.00	0.04	−0.02	−0.09	−0.04	0.01	−0.00	0.09
0.06	0.07	−0.09	0.03	0.05	0.00	0.01	0.00	0.10
−0.04	−0.02	−0.00	0.00	−0.04	−0.02	0.01	−0.00	0.01
0.00	−0.02	0.03	−0.07	−0.06	−0.03	0.00	0.02	−0.08

如果能用物理方法记录f_t(如绪论中的例1，采用经纬仪照相法获得轨道真实值)，则可直接从z_t中减去ρ_t. 否则,常用的办法是从z_t中减去它的样本均值$\hat{\mu}=\frac{1}{N}\sum_{k=1}^{N}z_k$，即用$w_t=z_t-\hat{\mu}$代替$z_t$进行时间序列分析. 关于估计量$\hat{\mu}$的一些统计性质，将在本章§4中做较详细的讨论.

现在,我们根据上述要领,对第一章所举的六个伪随机数列进行模型的初步识别,借此可以熟悉初步识别的方法. 在第一章§2中我们分别绘出了它们的一段样本坐标曲线，以及它们的样本自协方差函数曲线. 在第二章 §1 中，在有了随机模型的概念之后，我们便列举了构造这六个伪随机数列的模型(2.1.16a)～(2.1.16f). 在第二章§2里，又进一步算出了它们的理论自相关函数. 现在的问题是,假设我们并不知道构造这些数列的模型,如何根据它们的一段样本数据$z_1, z_2\cdots, z_N$ $(N=300)$，对它们的模型进行判断,也就是进行模型识别. 为节省篇幅,我们不一一列出这六组伪随机数序列，而只列出它们的样本自相关和样本偏相关函数$\hat{\rho}_k$和$\hat{\varphi}_{kk}$的若干数值(见表3.2.1和表3.2.2)，并把它们依次绘在本节图3.2.1和图3.2.2中. 为了进行对比，我们还把它们各自的理论自相关和理论偏相关函数ρ_k和φ_{kk}在相应的图中用虚线绘出.

下面,我们对这些实例逐个进行分析.

图3.2.1(a)的$\hat{\rho}_k$，当$k>1$时非常接近零，图3.2.2(a)的$\hat{\varphi}_{kk}$则显现出拖尾的趋势,因此大致可定出$q=1$, $p=0$. 若要仔细考察时，按要领2所叙述的方法，先计算出$\mathrm{var}(\hat{\rho}_k)\approx\frac{1+2\hat{\rho}_1^2}{300}$，然后检查$\hat{\rho}_2, \hat{\rho}_3, \cdots, \hat{\rho}_{30}$$\left(我们取\ M=\frac{N}{10}\right)$中满足$|\hat{\rho}_k|<\left(\frac{1+2\hat{\rho}_1^2}{300}\right)^{\frac{1}{2}}$的比例是否达到68.3%，其结果是$\mathrm{var}(\hat{\rho}_k)\approx 0.065$，而上述比例是$\frac{27}{39}=69.2\%$. 因此$q=1$这一判断可以接受. 于是初步识别$z_t$为MA(0, 1)序列,这与真实模型相符.

图 3.2.1 (b) 的 $\hat{\rho}_k$ 有截尾趋势，而 $\hat{\rho}_{kk}$ 有拖尾趋势，初看起来，也可以定 z_t 为 MA(0, 1)型. 但是根据第二章 § 2 例 2，MA(0,1) 的允许域为 $|\rho_1|<\frac{1}{2}$，而此处 $\hat{\rho}_1=-0.59$ 不在允许域内，因此应当考虑 MA(0, 2) 模型. 注意 $\hat{\rho}_2=0.10$，根据第二章 § 2 例 3，MA(0, 2) 的允许域条件检验如下:

$$\hat{\rho}_2+\hat{\rho}_1=-0.49>-\frac{1}{2}$$

$$\hat{\rho}_2-\hat{\rho}_1=0.69>-\frac{1}{2}, \qquad \hat{\rho}_2<\frac{1}{6}$$

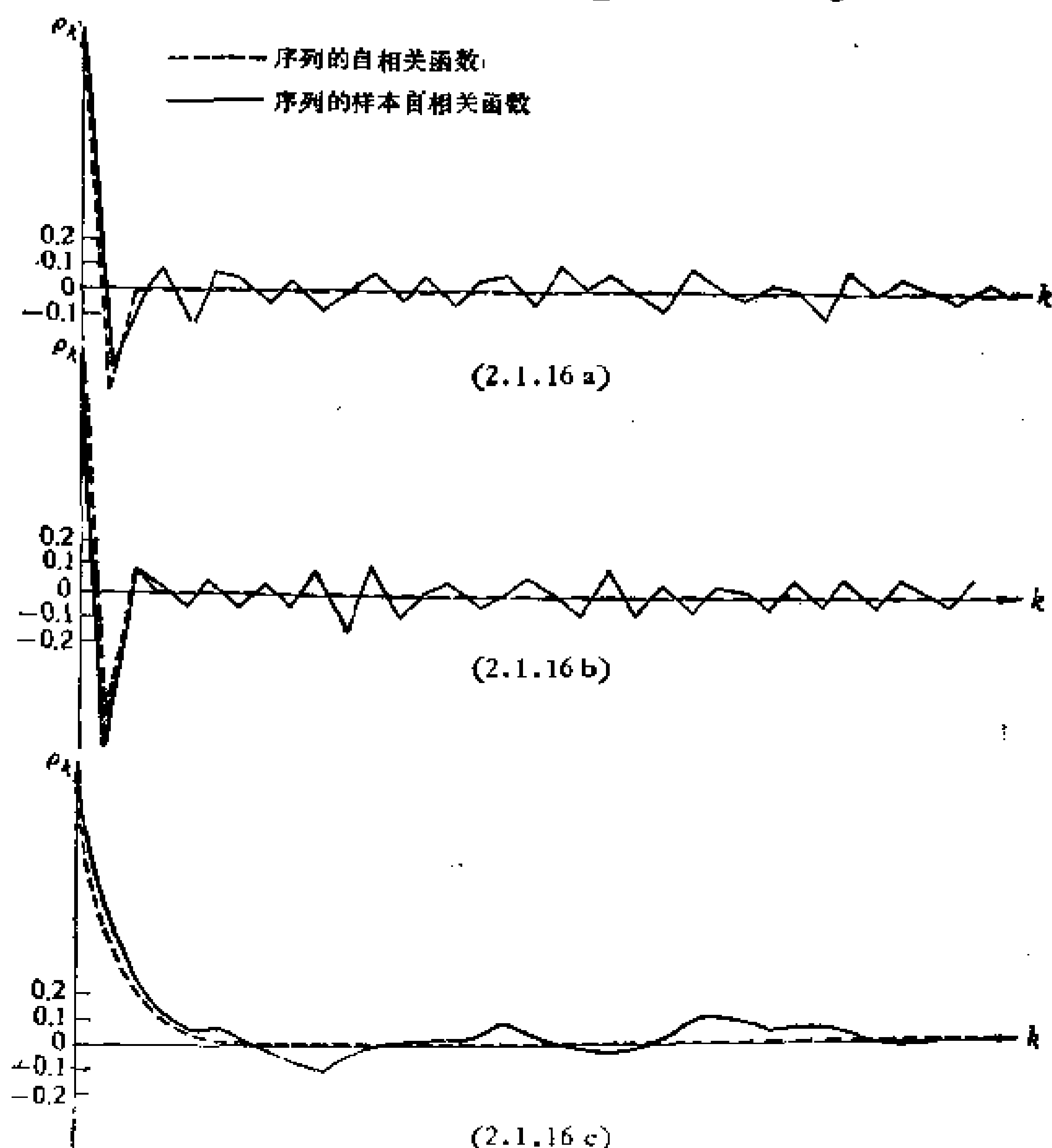
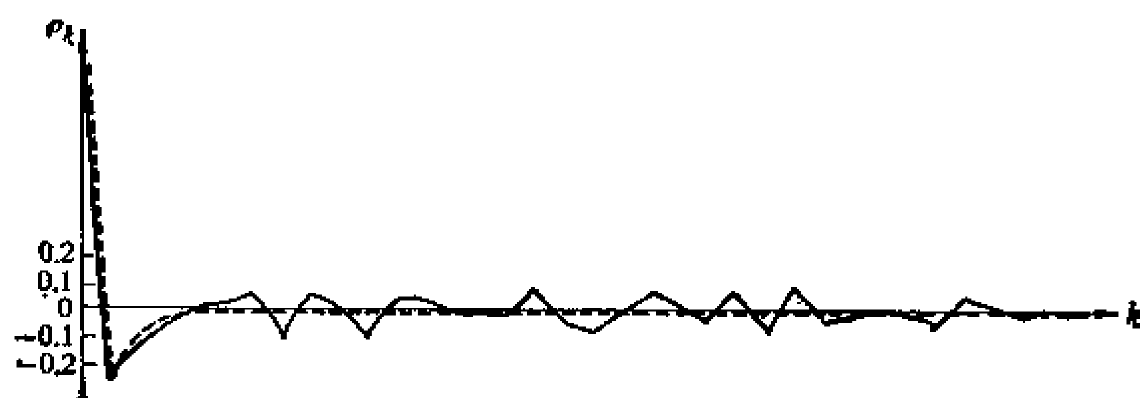


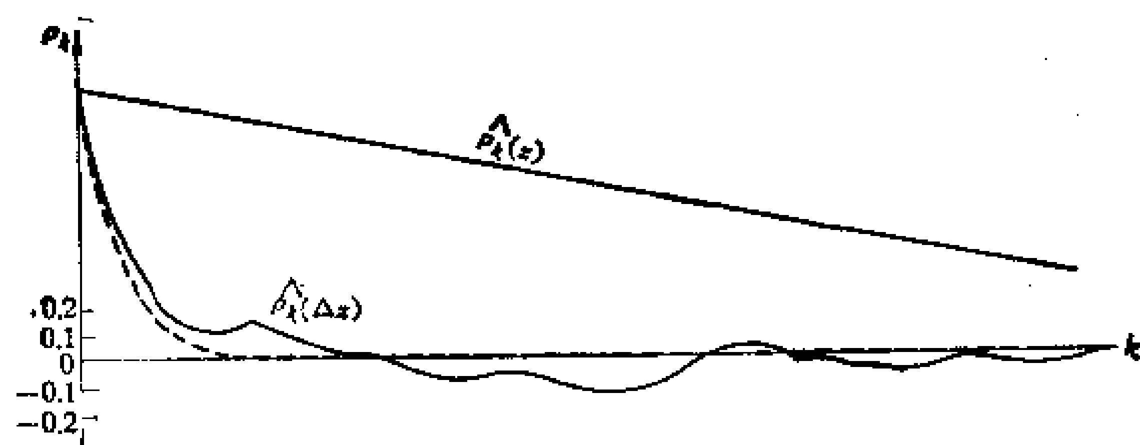
图 3.2.1 (2.1.16) 诸序列的自相关与样本自相关函数



(2.1.16 d)



(2.1.16 e)



(2.1.16 f) (2.1.16 f) 差分

图 3.2.1 (续)

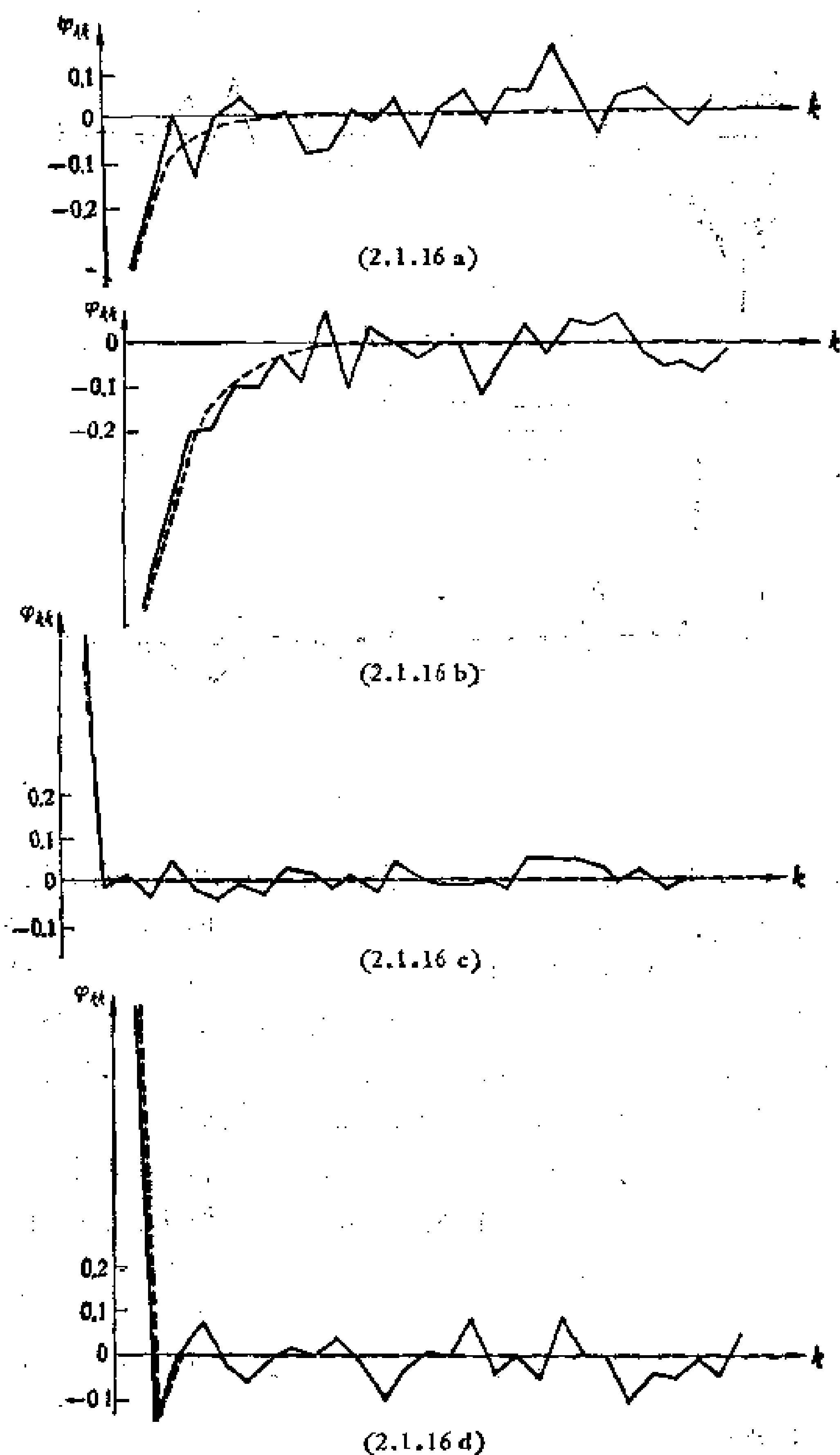
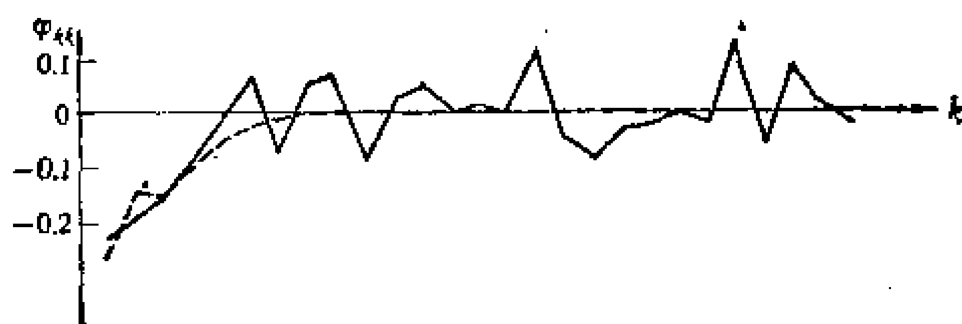
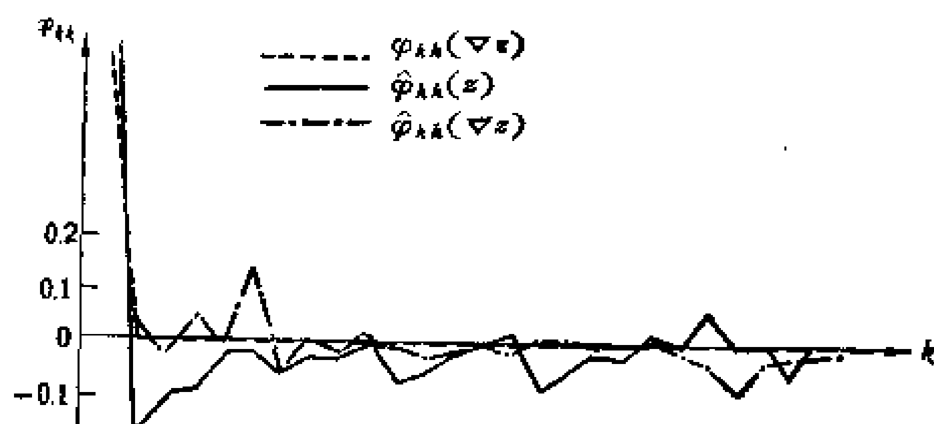


图 3.2.2 (2.1.16) 诸序列的偏相关与样本偏相关函数



(2.1.16 e)



(2.1.16 f)

图 3.2.2 (续)

这说明 $(\hat{\rho}_1, \hat{\rho}_2)$ 在此域内. 为了仔细审查, 仍可按要领 2 进行检验,此处从略. 当然, 也可以继续检查用 MA(0, 3) 模型是否可行, 这就是下一节将要讨论的识别的多样性问题. 一般在应用时都遵循以下两点: 其一,尽量采用低阶模型; 其二, 尽量使模型系数的估计不要接近平稳可逆域的边界. 由上面的讨论, 我们将此数值序列 z_t 定为 MA(0, 2) 序列,这也是与真实情况符合的.

图 3.2.1 (c) 的 $\hat{\rho}_k$ 明显地拖尾, 而 $\hat{\varphi}_{kk}$ 近于截尾, 因此可暂定 $p=1$, $q=0$. 为要仔细考察,利用 (3.1.13) 式和 $\hat{\varphi}_{kk}$ 的渐近正态性,发现 $\hat{\varphi}_{22}, \hat{\varphi}_{33}, \cdots, \hat{\varphi}_{28,28}$ 中满足 $|\hat{\varphi}_{kk}| < \frac{1}{\sqrt{300}}$ 的比例为 $\frac{21}{26} \approx 80\%$, 即可以接受上述判断. 于是初步识别为 AR(1, 0) 序列,与真实情况相符.

图 3.2.1 (d) 的 $\hat{\rho}_k$ 明显地拖尾，而 $\hat{\varphi}_{kk}$ 近于截尾，如果暂定 $p=1$, $q=0$，经与上例类似的考察手续后，发现有 70% 的 $\hat{\varphi}_{kk}$ 满足 $|\hat{\varphi}_{kk}|<\frac{1}{\sqrt{300}}$，因此可以接受. 但是注意到 $\hat{\varphi}_{22}=-0.15$，若认定模型为 AR(1, 0)，则 φ_{22} 理应为零，这样 $\hat{\varphi}_{22}$ 与零之差超过了 $E\hat{\varphi}_{22}^2=\frac{1}{\sqrt{300}}$ 的两倍多，所以在很大程度上可能 $\varphi_{22}\neq 0$，也就是说，值得考虑 AR(2, 0) 模型. 若改用这一模型再检查 $|\hat{\varphi}_{kk}|<\frac{1}{\sqrt{300}}$ 的比例时，结果约为 74%，同样可以接受. 通过这个例子，我们可以得到两点启发：其一，初步识别常常出现多样性，其二，在识别取为 AR($p-1$, 0) 和 AR(p, 0) 都可接受时，特别要注意比较 $|\hat{\varphi}_{pp}|$ 与 $\frac{1}{\sqrt{N}}$. 如果 $|\hat{\varphi}_{pp}|\ll\frac{1}{\sqrt{N}}$，可认为 $\varphi_{pp}=0$，即判断 z_t 为 AR($p-1$, 0) 序列；如果 $\hat{\varphi}_{pp}>\frac{2}{\sqrt{N}}$ $\left(\text{甚至于}>\frac{3}{\sqrt{N}}\right)$，则应否定 $\varphi_{pp}=0$，即判断 z_t 为 AR(p, 0) 序列；若 $\hat{\varphi}_{pp}$ 接近 $\frac{1}{\sqrt{N}}$，则两种判断都可以采用. 当然最后还可以通过第五章的方法鉴定哪一种更为合适. 对于 MA 模型的识别，类似的考虑也适用，不过在处理图 3.2.1 (b) 时我们已经看到，此时允许域也常常能帮助我们判断哪种模型更合理.

图 3.2.1 (e) 的 $\hat{\rho}_k$ 和图 3.2.2 (e) 的 $\hat{\varphi}_{kk}$ 都有明显的拖尾形状，根据要领 4 所述，这时不象截尾那样容易确定阶数. 从原则上说，只有从低阶向高阶逐个试验才能作出判断. 当然，利用 $\hat{\rho}_k$ 和 $\hat{\varphi}_{kk}$ 的渐近性质，及对不同的 p, q, ρ_k 和 φ_{kk} 的不同性质，也可以摸索出一些粗略估计 p, q 的办法，不过一般说来，这类办法既烦琐而效果又不显著. 比如，在本例中，由于 $\hat{\rho}_2:\hat{\rho}_1\approx 0.56$, $\hat{\rho}_3:\hat{\rho}_2\approx 0.46$ 可以认为在 $k>1$ 后 $\rho_k:\rho_{k-1}$ 近似为 0.4～0.5 之间的常数，也就是猜想 $\rho_k=\alpha\varphi_1^k (k\geq 1)$. 根据 (2.2.17) 式，这意味着 $p=1$,

和 $1>q-p=q-1$，即 $q=1$. 为什么只考察 $\hat{\rho}_1, \hat{\rho}_2, \hat{\rho}_3$ 之间的近似等比关系呢？ 因为，若真实的 $\rho_k=\alpha\varphi_1^k$ 而 φ_1 在 0.4 与 0.5 之间的话，则 $\rho_k \leqslant \alpha(0.5)^k$，于是从 $k=4$ 起有 $\rho_k<0.063\alpha$，但 $(E\tilde{\rho}_k^2)^{\frac{1}{2}} \approx 0.07$（根据（3.1.8）式计算），再注意，可以表 $\hat{\rho}_k:\hat{\rho}_{k-1}=\frac{\tilde{\rho}_k+\alpha\varphi_1^k}{\tilde{\rho}_{k-1}+\alpha\varphi_1^{k-1}}$. 可见当 $k \geqslant 4$ 时，在这个比值中 $\rho_k:\rho_{k-1}=\frac{\varphi_1^k}{\varphi_1^{k-1}}=\varphi_1$ 几乎完全被 $\tilde{\rho}_k$ 的随机性掩盖了. 这也正是上述方法的弱点所在.

图 3.2.2 (f) 的 $\hat{\varphi}_{kk}$ 虽然类似截尾，但是 ρ_k 不仅不截尾，而且也看不出有趋于零的迹象，所以我们只能认为它既不截尾也不拖尾. 根据要领 3，在 $\hat{\rho}_k, \hat{\varphi}_{kk}$ 中，只要有一个呈现出既不截尾也不拖尾的现象，就应依次考查 $\nabla z_k, \nabla^2 z_k$ 等差分序列的自相关与偏相关函数. 首先，我们对 ∇z_k 进行分析，并将 ∇z_k 的自相关与偏相关函数绘在图 3.2.2 中. 分析结果与图 3.2.1 (c) 相似，故判断 ∇z_t 为 AR(1, 0) 模型. 这样便定出 $d=1, p=1, q=0$，即初步识别后定 z_t 为 ARIMA (1, 1, 0) 序列.

§3. 参数的相关矩估计方法

利用前一节的方法，对于一串样本数据列 $z_1, z_2, \cdots, z_N$，经过分析后，可以初步识别原随机序列 z_t 的类型，即确定 ARIMA 序列的阶数 p, d, q. 在这一节里，我们将介绍如何利用样本自协方差函数或样本自相关函数，对 z_t 的模型参数进行粗略估计，所谓"粗"估计是相对下一节的"精"估计而言的. 这种估计的思想十分简单，它类似于数理统计中采用的矩估计方法. 假定 z_t 是 ARIMA(p, d, q) 序列，即 $w_t=\nabla^d z_t$ 是 ARMA 序列，根据第二章的内容，w_t 的自协方差或自相关函数可以通过 w_t 的模型参数 $\boldsymbol{\varphi}, \boldsymbol{\theta}$ 和 σ_a^2 表达出来. 所有用到的公式都已在第二章 §2 中给出，只要将这些公式中的 γ_k 或 ρ_k 换成 §1 中的 $\hat{\gamma}_k$ 或 $\hat{\rho}_k$，从中解出的 $\hat{\boldsymbol{\varphi}}=(\hat{\varphi}_1, \hat{\varphi}_2, \cdots, \hat{\varphi}_p)^\tau$, $\hat{\boldsymbol{\theta}}=(\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_q)^\tau$ 和 $\hat{\sigma}_a^2$，便分别是 $\boldsymbol{\varphi}$, $\boldsymbol{\theta}$ 和 σ_a^2 的相关矩估计，或简称为矩估计. 由于 $\hat{\gamma}_k, \hat{\rho}_k$ 是 γ_k, ρ_k 的

估计，当N足够大以后，它们分别接近各自的真值，所以$\hat{\boldsymbol{\varphi}}$，$\hat{\boldsymbol{\theta}}$和$\hat{\sigma}_a^2$也可指望接近它们各自的真值．这种方法简单易懂，不过在某些情况下精度较差．但是当w_t为AR$(p, 0)$序列时，这种估计与下一章中的精估计相差无几．此外，在实际使用时，还须注意$\hat{\rho}_k$是否在允许域内，只有在允许域内时，所得到的参数矩估计才在相应的平稳域或可逆域内．

下面分别就三种不同模型叙述参数的矩估计法，其中包括了具体公式，求解方法和举例说明．

一、AR 模型参数的矩估计

在(2.2.11)式中，以$\hat{\rho}_k$代替ρ_k并解出$\hat{\varphi}_1, \hat{\varphi}_2, \cdots, \hat{\varphi}_p$，即得

$$\begin{pmatrix} \hat{\varphi}_1 \\ \hat{\varphi}_2 \\ \vdots \\ \hat{\varphi}_p \end{pmatrix} = \begin{pmatrix} 1 & \hat{\rho}_1 & \cdots & \hat{\rho}_{p-1} \\ \hat{\rho}_1 & 1 & \cdots & \hat{\rho}_{p-2} \\ \cdots & \cdots & \cdots & \cdots \\ \hat{\rho}_{p-1} & \hat{\rho}_{p-2} & \cdots & 1 \end{pmatrix}^{-1} \begin{pmatrix} \hat{\rho}_1 \\ \hat{\rho}_2 \\ \vdots \\ \hat{\rho}_p \end{pmatrix} \tag{3.3.1}$$

文献中常称这样解出的$\hat{\varphi}_1, \hat{\varphi}_2, \cdots, \hat{\varphi}_p$为$\varphi_1, \varphi_2, \cdots, \varphi_p$的Yule-Walker估计．再将这些$\hat{\varphi}_1, \hat{\varphi}_2, \cdots, \hat{\varphi}_p$代替(2.2.13)式中的$\varphi_1, \varphi_2, \cdots, \varphi_p$，便得到

$$\hat{\sigma}_a^2 = \hat{\gamma}_0 - \sum_{j=1}^{p} \hat{\varphi}_j \hat{\gamma}_j = \hat{\gamma}_0 - \sum_{i,j=1}^{p} \hat{\varphi}_i \hat{\varphi}_j \hat{\gamma}_{j-i} \tag{3.3.2}$$

(3.3.1)和(3.3.2)式就是AR模型参数矩估计的全部公式，它们不仅简单易行，而且与下一章的AR模型的参数精估计的精度相差很少．下面以AR$(1, 0)$和AR$(2, 0)$为例，说明这些公式的具体应用情况．

AR$(1, 0)$模型参数的矩估计为

$$\hat{\varphi}_1 = \hat{\rho}_1 = \frac{\hat{\gamma}_1}{\hat{\gamma}_0}, \qquad \hat{\sigma}_a^2 = \hat{\gamma}_0 - \hat{\gamma}_1 \hat{\varphi}_1 = \hat{\gamma}_0 (1 - \hat{\rho}_1)^2 \tag{3.3.3}$$

AR$(2, 0)$模型参数的矩估计为

$$\begin{gathered} \hat{\varphi}_1 = \frac{\hat{\rho}_1 (1 - \hat{\rho}_2)}{1 - \hat{\rho}_1^2}, \qquad \hat{\varphi}_2 = \frac{\hat{\rho}_2 - \hat{\rho}_1^2}{1 - \hat{\rho}_1^2} \\ \hat{\sigma}_a^2 = \hat{\gamma}_0 (1 - \hat{\varphi}_1 \hat{\rho}_1 - \hat{\varphi}_2 \hat{\rho}_2) \end{gathered} \tag{3.3.4}$$

利用以上二式，我们就上一节的数值实例 (c) 和 (d) 对模型 (2.1.16 c) 和 (2.1.16 d) 的参数求矩估计，结果如下 ($\hat{\rho}_k$ 的具体数值见本章 §2 表 3.2.1)：

(2.1.16 c) 模型参数的矩估计为：$\hat{\varphi}_1 = 0.56$, $\hat{\sigma}_a^2 = 1.06$, 与它们的真值 $\varphi_1 = 0.5$, $\sigma_a^2 = 1$ 相比，误差分别为 $\tilde{\varphi}_1 = 0.06$, $\tilde{\sigma}_a^2 = 0.06$，根据附录 §4 和 §5 可知

$$E\tilde{\varphi}_1^2 \approx \frac{1}{N}(1-\varphi_1^2) = \frac{3}{300\times 4} = \frac{1}{400}$$

$$E(\tilde{\sigma}_a^2)^2 \approx \frac{2\sigma_a^4}{N} = \frac{2}{300} = \frac{1}{150}$$

由此可见，$\tilde{\varphi}_1 = 0.06$ 与 $\sqrt{E\tilde{\varphi}_1^2} \approx \frac{1}{20} = 0.05$ 相近；$\tilde{\sigma}_a^2 = 0.06$ 与 $\sqrt{E(\tilde{\sigma}_a^2)^2} \approx \sqrt{\frac{1}{150}} \cong 0.08$ 也比较相近.

(2.1.16 d) 模型参数的矩估计为：$\hat{\varphi}_1 = 0.91$, $\hat{\varphi}_2 = -0.14$ $\hat{\sigma}_a^2 = 0.96$ 与它们的真值 $\varphi_1 = 1$, $\varphi_2 = -0.24$, $\sigma_a^2 = 1$ 相比，误差分别为 $\tilde{\varphi}_1 = -0.09$, $\tilde{\varphi}_2 = 0.10$, $\tilde{\sigma}_a^2 = -0.04$. 仍然根据附录 §4 的结果可知

$$E\tilde{\varphi}_1^2 = E\tilde{\varphi}_2^2 \approx \frac{1}{N}(1-\varphi_2^2) \cong 0.31$$

$$E(\tilde{\sigma}_a^2)^2 \approx \frac{2\sigma_a^4}{N} = \frac{1}{150}$$

由此可见，$\sqrt{E\tilde{\varphi}_1^2} = \sqrt{E\tilde{\varphi}_2^2} \cong 0.06$, $\sqrt{E(\tilde{\sigma}_a^2)^2} \cong 0.08$. 虽然 $|\tilde{\varphi}_1|$, $|\tilde{\varphi}_2|$ 比它们的均方根值稍大，但都没有超过 2 倍.

二、MA 模型参数的矩估计

在 (2.2.3) 式中，以 $\hat{\gamma}_k$ 代替 γ_k 便得到方程组

$$\hat{\gamma}_k = \begin{cases} \hat{\sigma}_a^2(1+\hat{\theta}_1^2+\hat{\theta}_2^2+\cdots+\hat{\theta}_q^2), & k=0 \\ \hat{\sigma}_a^2(-\hat{\theta}_k+\hat{\theta}_1\hat{\theta}_{k+1}+\cdots+\hat{\theta}_{q-k}\hat{\theta}_q), & 1\leqslant k\leqslant q \end{cases} \tag{3.3.5}$$

由此式解出的 $\hat{\sigma}_a^2, \hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_q$ 便是 MA 模型参数的矩估计. (3.3.5) 式是一个 $q+1$ 元二次方程组，下面给出它的三种解法：

1. 直接解法：当 $q=1$ 时，(3.3.5) 式为

$$\begin{cases}\hat{\gamma}_0=\hat{\sigma}_a^2(1+\hat{\theta}_1^2)\\ \hat{\gamma}_1=-\hat{\sigma}_a^2\hat{\theta}_1\end{cases}$$

由后一式解得 $\hat{\theta}_1=-\frac{\hat{\gamma}_1}{\hat{\sigma}_a^2}$，代入前一式得到

$$\hat{\sigma}_a^4-\hat{\gamma}_0\hat{\sigma}_a^2+\hat{\gamma}_1^2=0$$

由此解出 $\hat{\sigma}_a^2$，再代入 $\hat{\theta}_1$ 便知

$$\hat{\sigma}_a^2=\hat{\gamma}_0\frac{1\pm\sqrt{1-4\hat{\rho}_1^2}}{2},\quad \hat{\theta}_1=\frac{-2\hat{\rho}_1}{1\pm\sqrt{1-4\hat{\rho}_1^2}}$$

至于在 $\hat{\sigma}_a^2$ 和 $\hat{\theta}_1$ 的两种可能值中应取哪一个，由可逆性条件帮助判断。由第二章 §1 例 5 MA(0, 1) 的可逆域推知，应当取满足条件 $|\hat{\theta}_1|<1$ 作为 θ_1 的估计。

注意由

$$\frac{2\hat{\rho}_1}{1+\sqrt{1-4\hat{\rho}_1^2}}\cdot\frac{2\hat{\rho}_1}{1-\sqrt{1-4\hat{\rho}_1^2}}=1$$

可得

$$\left|\frac{2\hat{\rho}_1}{1+\sqrt{1-4\hat{\rho}_1^2}}\right|<1$$

因此应取 σ_a^2 和 θ_1 的矩估计为

$$\hat{\sigma}_a^2=\hat{\gamma}_0\frac{1+\sqrt{1-4\hat{\rho}_1^2}}{2},\quad \hat{\theta}_1=\frac{-2\hat{\rho}_1}{1+\sqrt{1-4\hat{\rho}_1^2}}\qquad(3.3.6)$$

当 $q=2$ 时，(3.3.5) 式为

$$\begin{cases}\hat{\gamma}_0=\hat{\sigma}_a^2(1+\hat{\theta}_1^2+\hat{\theta}_2^2)\\ \hat{\gamma}_1=\hat{\sigma}_a^2(-\hat{\theta}_1+\hat{\theta}_1\hat{\theta}_2)\\ \hat{\gamma}_2=-\hat{\sigma}_a^2\hat{\theta}_2\end{cases}$$

由此方程的后两式解出

$$\hat{\theta}_2=-\frac{\hat{\gamma}_2}{\hat{\sigma}_a^2},\qquad \hat{\theta}_1=-\frac{\hat{\gamma}_1}{\hat{\sigma}_a^2+\hat{\gamma}_2}$$

将它们代入方程的第一式可得

$$\hat{\gamma}_0 = \hat{\sigma}_a^2\left\{1 + \frac{\hat{\gamma}_2^2}{\hat{\sigma}_a^4} + \frac{\hat{\gamma}_1^2}{(\hat{\sigma}_a^2 + \hat{\gamma}_2)^2}\right\}$$

这是 $\hat{\sigma}_a^2$ 的四次方程，它有四个根，因此 $\hat{\theta}_1$，$\hat{\theta}_2$ 也有四种可能的解，根据可逆性条件，从中可以选出唯一的合理解来. 事实上，由第二章 § 1 例 6 给出的 MA(0, 2) 的可逆域，经过初等但烦琐的讨论，可以解出 θ_2，θ_1，σ_a^2 的矩估计分别为

$$\hat{\theta}_2 = \frac{1}{2} - \frac{1}{4\hat{\rho}_2} - \frac{1}{2\hat{\rho}_2}\sqrt{\left(\hat{\rho}_2 + \frac{1}{2}\right)^2 - \hat{\rho}_1^2}$$

$$\pm\sqrt{\left(\frac{1}{2} - \frac{1}{4\hat{\rho}_2} - \frac{1}{2\hat{\rho}_2}\sqrt{\left(\hat{\rho}_2 + \frac{1}{2}\right)^2 - \hat{\rho}_1^2}\right)^2 - 1}$$

式中的正负号“±”依 $\hat{\rho}_2 > 0$ 或 < 0 而分别取 + 或 −，

$$\hat{\theta}_1 = \frac{\hat{\rho}_1\hat{\theta}_2}{\hat{\rho}_2(1 - \hat{\theta}_2)}, \qquad \hat{\sigma}_a^2 = -\hat{\gamma}_0\frac{\hat{\rho}_2}{\hat{\theta}_2}$$

按上面的方法从 θ_q 到 θ_1 逐步消元解 (3.3.5) 式，最后需要解 $\hat{\sigma}_a^2$ 的 $2q$ 次方程，而在 $q \geqslant 3$ 以后，这类方程一般只能用数值解法. 所以只有当 q 比较小时，或者高次方程易解时，才用直接解法. 当 q 较大时，可以用下面的迭代计算方法.

2. 线性迭代法：先将 (3.3.5) 式改写成

$$\begin{cases}\hat{\sigma}_a^2 = \hat{\gamma}_0(1 + \hat{\theta}_1^2 + \cdots + \hat{\theta}_q^2)^{-1} \\ \hat{\theta}_k = -\left(\dfrac{\hat{\gamma}_k}{\hat{\sigma}_a^2} - \hat{\theta}_1\hat{\theta}_{k+1} - \cdots - \hat{\theta}_{q-k}\hat{\theta}_q\right), \ k = 1, 2, \cdots, q\end{cases} \tag{3.3.7}$$

给出 $\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_q$ 和 $\hat{\sigma}_a^2$ 的一组初值（例如 $\hat{\boldsymbol{\theta}} = 0$，$\hat{\sigma}_a^2 = \hat{\gamma}_0$），代入 (3.3.7) 式右边，由此得到等式左边的值 $\hat{\sigma}_a^2(1)$ 和 $\hat{\boldsymbol{\theta}}(1) = (\hat{\theta}_1(1), \hat{\theta}_2(1), \cdots \hat{\theta}_q(1))^\tau$，称为 $\hat{\sigma}_a^2$ 和 $\hat{\boldsymbol{\theta}}$ 的一步迭代值. 再将它们代入 (3.3.7) 式右边又得到 $\hat{\sigma}_a^2(2)$ 和 $\hat{\boldsymbol{\theta}}(2)$. 以此方式重复迭代下去，直到某个 $\hat{\sigma}_a^2(m)$ 和 $\hat{\boldsymbol{\theta}}(m)$ 与 $\hat{\sigma}_a(m-1)$ 和 $\hat{\theta}(m-1)$ 的出入不大时（比如这两步迭代值之差的各分量绝对值都比预定的精度还小），便停止迭代. $\hat{\boldsymbol{\theta}}(m)$ 和 $\hat{\sigma}_a^2(m)$ 便可做为 (3.3.7) 式（即 (3.3.5) 式）的近似解.

3. Newton-Raphson 算法：将 (3.3.5) 式改写成

$$\begin{cases}\hat{\gamma}_0 = \hat{\sigma}_a^2 + (-\theta_1\hat{\sigma}_a)^2 + \cdots + (-\theta_q\hat{\sigma}_a)^2 \\ \hat{\gamma}_1 = \hat{\sigma}_a(-\hat{\theta}_1\hat{\sigma}_a) + (-\hat{\theta}_1\hat{\sigma}_a)(-\hat{\theta}_2\hat{\sigma}_a) + \cdots + (-\hat{\theta}_{q-1}\hat{\sigma}_a)(-\hat{\theta}_q\hat{\sigma}_a) \\ \cdots\cdots\cdots\cdots \\ \hat{\gamma}_q = \hat{\sigma}_a(-\hat{\theta}_q\hat{\sigma}_a)\end{cases} \tag{3.3.8}$$

我们暂时令 $\hat{\sigma}_k = -\hat{\sigma}_a\hat{\theta}_k\,(1 \leqslant k \leqslant q)$ 和 $\hat{\sigma}_0 = \hat{\sigma}_a$，于是 (3.3.8) 式又可写为

$$\begin{cases}\hat{\sigma}_0^2 + \hat{\sigma}_1^2 + \cdots + \hat{\sigma}_q^2 - \hat{\gamma}_0 = 0 \\ \hat{\sigma}_0\hat{\sigma}_1 + \hat{\sigma}_1\hat{\sigma}_2 + \cdots + \hat{\sigma}_{q-1}\hat{\sigma}_q - \hat{\gamma}_1 = 0 \\ \cdots\cdots\cdots\cdots \\ \hat{\sigma}_0\hat{\sigma}_q - \hat{\gamma}_q = 0\end{cases} \tag{3.3.9}$$

将上方程组左边各式分别记做 $\hat{f}_k = \hat{f}_k(\hat{\sigma}_0, \hat{\sigma}_1, \cdots, \hat{\sigma}_q)$ $(k = 0, 1, 2, \cdots, q)$，并记

$$\hat{\boldsymbol{\sigma}} = \begin{pmatrix}\hat{\sigma}_0 \\ \hat{\sigma}_1 \\ \vdots \\ \hat{\sigma}_q\end{pmatrix}, \quad \hat{\boldsymbol{f}} = \begin{pmatrix}\hat{f}_0 \\ \hat{f}_1 \\ \vdots \\ \hat{f}_q\end{pmatrix}, \quad F = \frac{\partial \hat{\boldsymbol{f}}}{\partial \hat{\boldsymbol{\sigma}}^\tau} = \begin{pmatrix}\frac{\partial \hat{f}_0}{\partial \hat{\sigma}_0} & \cdots & \frac{\partial \hat{f}_0}{\partial \hat{\sigma}_q} \\ \cdots & \cdots & \cdots \\ \frac{\partial \hat{f}_q}{\partial \hat{\sigma}_0} & \cdots & \frac{\partial \hat{f}_q}{\partial \hat{\sigma}_q}\end{pmatrix}$$

易见

$$F = \begin{pmatrix}\hat{\sigma}_0 & \hat{\sigma}_1 & \cdots & \hat{\sigma}_q \\ \hat{\sigma}_1 & \hat{\sigma}_2 & \cdots \ \hat{\sigma}_q & \\ & \cdots\cdots & & \\ \hat{\sigma}_q & & & \end{pmatrix} + \begin{pmatrix}\hat{\sigma}_0 & \hat{\sigma}_1 & \cdots & \hat{\sigma}_q \\ & \hat{\sigma}_0 & \cdots & \hat{\sigma}_{q-1} \\ & & & \cdots\cdots \\ & & & \hat{\sigma}_0\end{pmatrix}$$

按照 Newton-Raphson 迭代原则，如果第 i 步迭代值为 $\hat{\boldsymbol{\sigma}}(i)$，那么第 $i+1$ 步的 $\hat{\boldsymbol{\sigma}}(i+1)$ 须满足

$$\hat{\boldsymbol{f}}(i) + F(i)[\hat{\boldsymbol{\sigma}}(i+1) - \hat{\boldsymbol{\sigma}}(i)] = 0$$

即

$$\hat{\boldsymbol{\sigma}}(i+1) = \hat{\boldsymbol{\sigma}}(i) - F^{-1}(i)\hat{\boldsymbol{f}}(i) \tag{3.3.10}$$

只要给出初值 $\hat{\boldsymbol{\sigma}}(0)$，便可依 (3.3.10) 式进行迭代计算；并且按照前述同样办法决定何时停止迭代，以最后的 $\hat{\boldsymbol{\sigma}}(m)$ 做为 (3.3.9) 式的近似解. 于是 (3.3.5) 式的近似解取为 $\hat{\sigma}_a^2 = \hat{\sigma}_0^2(m)$，

$$\hat{\theta}_k = -\frac{\hat{\sigma}_k(m)}{\hat{\sigma}_0(m)} \quad (1 \leqslant k \leqslant q)$$

利用所介绍的各种算法，对模型 (2.1.16 a) 和 (2.1.16 b) 的参

数求矩估计（本章§2图3.1.1(a)、图3.1.1(b)与表3.2.1(a)、表3.2.1(b)列出了z_t的样本自相关函数值），其结果如下：

(2.1.16a)的模型参数矩估计为

$$\hat{\theta}_1=\frac{-2\hat{\rho}_1}{1+\sqrt{1-4\hat{\rho}_1^2}}=0.39,\quad \hat{\sigma}_a^2=\hat{\gamma}_0+\frac{1+\sqrt{1-4\hat{\rho}_1^2}}{2}=1.06$$

它们与各自的真值 $\theta_1=0.5$，$\sigma_a^2=1$ 的误差分别为 $\tilde{\theta}_1=-0.11$，$\tilde{\sigma}_a^2=0.06$。

(2.1.16 b)的模型参数矩估计 $\hat{\theta}_1, \hat{\theta}_2$ 和 $\hat{\sigma}_a^2$，按方法2进行计算，其初始值可取为 $\hat{\boldsymbol{\theta}}(0)=(1,0)^\tau$，$\hat{\sigma}_a^2(0)=\frac{\hat{\gamma}_0}{2}$. 这是因为根据本章§2对图3.2.1(b)的分析，它近似于MA(0, 1)模型，只是由于$\hat{\rho}_1$不在允许域内，才改定为MA(0, 2)模型，所以我们不妨就从MA(0, 1)模型可逆域的边界上开始迭代. 头几步迭代的具体结果如下：

$$\hat{\boldsymbol{\theta}}(0)=(1,\ 0)^\tau,\ \hat{\sigma}_a^2(\sigma)=1.1$$
$$\hat{\boldsymbol{\theta}}(1)=(1.1,\ -0.24)^\tau,\quad \hat{\sigma}_a^2(1)=0.96$$
$$\hat{\boldsymbol{\theta}}(2)=(1.1,\ -0.23)^\tau,\quad \hat{\sigma}_a^2(2)=0.98$$

可见只迭代两步，诸参数的变动就不太大了，若继续迭代下去也无多少改进. 于是可取矩估计近似为 $\hat{\theta}_1=1.1$，$\hat{\theta}_2=-0.23$，$\hat{\sigma}_a^2=0.98$，它们与各自的真实值 $\theta_1=1$，$\theta_2=-0.24$，$\sigma_a^2=1$ 的误差分别是 $\tilde{\theta}_1=0.1$，$\tilde{\theta}_2=0.01$，$\tilde{\sigma}_a^2=-0.02$。

三、ARMA 模型参数的矩估计

为了给出 ARMA(p, q) 模型参数的矩估计，我们分三步进行.

1. 首先给出自回归参数向量$\boldsymbol{\varphi}$的矩估计：在(2.2.18)式中，以$\hat{\rho}_k$代替ρ_k可以解出 $\hat{\boldsymbol{\varphi}}^\tau=(\hat{\varphi}_1, \hat{\varphi}_2, \cdots, \hat{\varphi}_p)$，即

$$\begin{pmatrix}\hat{\varphi}_1\\ \hat{\varphi}_2\\ \vdots\\ \hat{\varphi}_p\end{pmatrix}=\begin{pmatrix}\hat{\rho}_q & \hat{\rho}_{q-1} & \cdots & \hat{\rho}_{q-p+1}\\ \hat{\rho}_{q+1} & \hat{\rho}_q & \cdots & \hat{\rho}_{q-p+2}\\ \cdots & \cdots & \cdots & \cdots\\ \hat{\rho}_{q+p-1} & \hat{\rho}_{q+p-2} & \cdots & \hat{\rho}_q\end{pmatrix}^{-1}\begin{pmatrix}\hat{\rho}_{q+1}\\ \hat{\rho}_{q+2}\\ \vdots\\ \hat{\rho}_{q+p}\end{pmatrix}\qquad(3.3.11)$$

2. 令

$$\bar{w}_t = w_t - \hat{\varphi}_1 w_{t-1} - \cdots - \hat{\varphi}_p w_{t-p} \tag{3.3.12}$$

其协方差函数为

$$\gamma_k(\bar{w}) = E(\bar{w}_t \bar{w}_{t+k}) = \sum_{i,j=0}^{p} \hat{\varphi}_i \hat{\varphi}_j \gamma_{k+j-i}$$

它类似于 (2.2.19) 式，其中记 $\hat{\varphi}_0 = -1$. 再以

$$\hat{\gamma}_k(\bar{w}) = \sum_{i,j=0}^{p} \hat{\varphi}_i \hat{\varphi}_j \hat{\gamma}_{k+j-i} \tag{3.3.13}$$

作为 $\gamma_k(\bar{w})$ 的估计，即把 $\hat{\gamma}_k(\bar{w})$ 当做 $\bar{w}_t$ 的样本自协方差函数.

3. 将 $\bar{w}_t$ 近似看作 MA(0, q) 序列，即

$$\bar{w}_t \cong a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}$$

并用前面介绍过的方法求 $\theta_1, \theta_2, \cdots \theta_q$ 和 σ_a^2 的矩估计. 首先列出方程

$$\begin{cases} \hat{\gamma}_0(\bar{w}) = \hat{\sigma}_a^2(1 + \hat{\theta}_1^2 + \cdots + \hat{\theta}_q^2) \\ \hat{\gamma}_k(\bar{w}) = \hat{\sigma}_a^2(-\hat{\theta}_k + \hat{\theta}_1\hat{\theta}_{k+1} + \cdots + \hat{\theta}_{q-k}\hat{\theta}_q), \quad k = 1, 2, \cdots, q \end{cases} \tag{3.3.14}$$

然后用直接解法或迭代法求这个方程组的解，其解就是 ARMA 模型的滑动平均参数 θ 和 σ_a^2 的矩估计. 由此可以想见，对 ARMA 模型而言，这种参数估计方法的精度更差.

在精度要求不太高时，还可以用图表法大致定出低阶模型的参数矩估计. 参考文献 [14] 中绘有 AR(2, 0) 模型 (ρ_1, ρ_2) 与 (φ_1, φ_2) 的对应关系曲线，MA(0, 2) 模型 (ρ_1, ρ_2) 与 (θ_1, θ_2) 的对应关系曲线，以及 ARMA (1, 1) 模型 (ρ_1, ρ_2) 与 (θ_1, φ_1) 的对应关系曲线.

利用这里所介绍的方法，我们对本章 § 2 中所讨论的数值例子 (e) 来求 (2.1.16 e) 模型参数的矩估计. 首先由 (3.3.11) 得到

$$\hat{\varphi}_1 = \frac{\hat{\rho}_2}{\hat{\rho}_1} = \frac{-0.13}{-0.23} = 0.56$$

再根据 (3.3.13) 式算得

$$\hat{\gamma}_0(\bar{w}) = (1 + \hat{\varphi}_1^2)\hat{\gamma}_0 - 2\hat{\varphi}_1\hat{\gamma}_1$$
$$= [1 + (0.56)^2]1.19 - 2 \times 0.56 \times 1.19 \times (-0.23) = 1.87$$
$$\hat{\gamma}_1(\bar{w}) = (1 + \hat{\varphi}_1^2)\hat{\gamma}_1 - \hat{\varphi}_1(\hat{\gamma}_0 + \hat{\gamma}_2)$$
$$= (1 + \hat{\varphi}_1^2)\hat{\gamma}_0\hat{\rho}_1 - \hat{\varphi}_1(1 + \hat{\rho}_2)\hat{\gamma}_0$$
$$= \hat{\gamma}_0[(1 + \hat{\varphi}_1^2)\hat{\rho}_1 - \hat{\varphi}_1 - \hat{\varphi}_1\hat{\rho}_2]$$
$$= 1.19\{[1 + (0.56)^2](-0.23) - 0.56 + 0.56 \times 0.23\} = -0.94$$

最后再利用 (3.3.6) 式便得到

$$\hat{\theta}_1 = \frac{-2\hat{\rho}_1(\bar{w})}{1 + \sqrt{1 - 4\hat{\rho}_1^2(\bar{w})}} = \frac{2 \times 0.5}{1 + \sqrt{1 - 4(0.5)^2}} = 1$$

$$\hat{\sigma}_a^2 = \hat{\gamma}_0(\bar{w})\frac{1 + \sqrt{1 - 4\hat{\rho}_1^2(\bar{w})}}{2} = 0.94$$

这些估计值与其真值 $\varphi_1 = 0.3$, $\theta_1 = 0.6$, $\sigma_a^2 = 1$ 相比,其误差分别为 $\tilde{\varphi}_1 = 0.26$, $\tilde{\theta}_1 = 0.4$, $\tilde{\sigma}_a^2 = -0.06$, 它们比较大，而且 $\hat{\theta}_1$ 在可逆域的边界上. 这正反映了矩估计是比较粗糙的.

一般说来,MA 和 ARMA 模型参数的矩估计较差，因此，对它们的参数,最好采用第四章的精估计方法进行估计. 最后,还有本章 § 2 中第六个数值例子 (f),它已被初步识别为 ARIMA(1, 1,0) 模型,其参数估计如下: 根据 ∇z_k 的样本自相关函数 (见本章 § 2 表 3.2.1(f)) 和 (3.3.3) 式可得

$$\hat{\varphi}_1 = \hat{\rho}_1(\nabla z) = 0.54$$
$$\hat{\sigma}_a^2 = \hat{\gamma}_0(\nabla z) - \hat{\rho}_1(\nabla z)\hat{\gamma}_1(\nabla z) = 0.94$$

与它们的真实值 $\varphi_1 = 0.5$, $\sigma_a^2 = 1$ 相比,其误差分别为 $\tilde{\varphi}_1 = 0.04$, $\tilde{\sigma}_a = -0.06$.

到此为止,我们已经根据第一章 § 2 所给出的六组数据列(见第一章 § 2 图 1.2.1), 建立了它们各自的模型, 并且与构造它们的真实模型 (2.1.16) 诸式进行了比较. 为了从这些数值例子中得到更多的启示, 我们把这六个例子的模型与参数估计情况列成下表:

表 3.3.1 （2.1.16）诸序列的初步识别与估计结果

代号 \ 模型阶数 \ 模型参数			φ_1			φ_2			θ_1			θ_2			σ_a^2		
(2.1.16)	估 计	真 实	$\hat{\varphi}_1$	φ_1	$\tilde{\varphi}_1$	$\hat{\varphi}_2$	φ_1	$\tilde{\varphi}_1$	$\hat{\theta}_1$	θ_1	$\tilde{\theta}_1$	$\hat{\theta}_2$	θ_2	$\tilde{\theta}_2$	$\hat{\sigma}_a^2$	σ_a^2	$\tilde{\sigma}_a^2$
a	(0,0,1)	(0,0,1)							0.39	0.50	−0.11				1.06	1.00	0.06
b	(0,0,2)	(0,0,2)							1.10	1.00	0.10	−0.23	−0.24	0.01	0.98	1.00	−0.02
c	(1,0,0)	(1,0,0)	0.56	0.50	0.06										1.06	1.00	0.06
d	(2,0,0)	(2,0,0)	0.91	1.00	−0.09	−0.14	−0.24	0.10							0.96	1.00	−0.04
e	(1,0,1)	(1,0,1)	0.56	0.30	0.26				1.00	0.60	0.40				0.94	1.00	−0.06
f	(1,1,0)	(1,1,0)	0.54	0.50	0.04										0.94	1.00	−0.06

表中所以能够给出 $\hat{\boldsymbol{\theta}}$, $\hat{\boldsymbol{\varphi}}$ 和 $\hat{\sigma}_a^2$ 的具体值，是因为它们的真实模型和参数是知道的. 对于实际随机现象中时间序列的样本序列，其真实模型和参数一般无法知道，因此，利用它们建立起模型和参数估计以后，无法与真实值进行比较. 为了了解这种估计的精度，只有通过它们的渐近性质来了解(见本章 §1 和附录 §5 的内容)，在那些渐近公式中（如 (3.1.7) 和 (3.1.8) 式等)，又包含了真实参数，最终须将所得估计值代替这些真实参数，才能对估计的渐近分布或方差有所了解. 尽管如此，这毕竟是了解估计方法的可靠程度的基本手段. 在统计学中，另一种考查估计方法优劣的办法，就是采用大量的模拟计算，直接了解估计量与其真值的误差情况. 我们所引用的这六个数值例子，就是按照这后一种办法进行的. 由于受篇幅所限，我们只用了六个典型模型的例子做为说明. 如果进行大量的同类模拟计算的参数估计，就会发现 AR 模型参数的矩估计一般精度较高，其它两种模型一般精度较差.

当样本长度 N 相同时，在通常情况下(即 $\boldsymbol{\varphi}$ 和 $\boldsymbol{\theta}$ 不在平稳可逆域的边界附近)，模型的总阶数 $p+q$ 越高，各参数的估计精度越差. 所以我们在本章 §2 中曾指出，应当尽量采用低阶模型作为初步识别. 这些规律有的将在第四章里重新指出. 除此而外，我们还可以用模拟方法来验证附录 §5 中关于 $\hat{\boldsymbol{\varphi}}$ 的渐近方差公式. 虽然表 3.3.1 所列的数值例子的数量不多，但是读者也能从中大致地看到本节的估计方法的可信程度，以及上述统计性规律的某些表现.

§4. 模型识别的多样性与均值的判断

从前面的例子可以看出，对一个样本序列进行模型识别时，有时可能给出几种不同的识别结果. 比如对模型 (2.1.16 d)，既可以识别为 AR(1, 0) 模型，也可以识别为 AR(2, 0) 模型. 当高阶参数 φ_p 或 θ_q 的值越接近于零时，越容易出现多样性的识别结果；当 $\boldsymbol{\varphi}$, $\boldsymbol{\theta}$ 接近平稳可逆域的边界时，也有这种情况出现. 下面举一例说明产生这种多样性的原因.

我们考虑以下的 ARMA(1,1) 模型

$$z_t - \varphi_1 z_{t-1} = a_t - \theta_1 a_{t-1} \tag{3.4.1}$$

其中 $\varphi_1 = 1 - \varepsilon$, $\varepsilon \cong 0$. 由此可以导出

$$\begin{aligned} z_t - z_{t-1} &= a_t - \theta_1 a_{t-1} - \varepsilon z_{t-1} \\ &= a_t - (\theta_1 + \varepsilon) a_{t-1} + \varepsilon(\theta_1 a_{t-2} - \varphi_1 z_{t-2}) \end{aligned} \tag{3.4.2}$$

根据第二章 § 2 例 6 可知

$$E z_{t-2}^2 = \gamma_0 = \frac{1 + \theta_1^2 - 2\varphi_1\theta_1}{1 - \varphi_1^2} \sigma_a^2 = \frac{1 + \theta_1^2 - 2\varphi_1\theta_1}{\varepsilon(2 - \varepsilon)} \ \sigma_a^2$$

而 (3.4.2) 式右边第三项不仅与前两项相互独立,而且其方差相对前两项是很小的,即

$$\begin{aligned} E\varepsilon^2(\theta_1 a_{t-2} - \varphi_1 z_{t-2})^2 &\leqslant 2\varepsilon^2[E(\theta_1 a_{t-2})^2 + \varphi_1^2 E z_{t-2}^2] \\ &= 2\varepsilon\left[\varepsilon\theta_1^2 + \varphi_1^2 \frac{1 + \theta_1^2 - 2\varphi_1\theta_1}{2 - \varepsilon}\right] \sigma_a^2 \end{aligned}$$

因此, (3.4.2) 式中的第三项几乎可以忽略. 这样 (3.4.2) 式与 ARIMA(0, 1, 1) 模型

$$z_t - z_{t-1} = a_t - (\theta_1 + \varepsilon) a_{t-1} \tag{3.4.3}$$

就很难区分了. 此外,根据样本自相关函数识别模型时,样本自相关函数又有一定的误差,因此, 在模型本身就很相近时, 识别时也就更容易出现多样性现象.

在出现多种可能的识别结果时, 一般说来, 识别结果中的 ARIMA (p, d, q) 模型阶数 $(p + d)$ 常常是相同的. 上面所举的例子,从理论模型上恰好说明了这一点, 因为 (3.4.1) 式为 ARMA (1, 0, 1) 模型,而它与 (3.4.3) 式十分相近,即与 ARIMA(0, 1, 1) 模型很相近,其中两个模型相应的 $(p + d)$ 都等于 1 .

虽然在识别中会出现多样性, 但是这些模型与真实模型还是十分相近的, 因为它们的自协方差函数非常接近. 我们已经指出过, 自协方差函数相同时, 相应的随机序列的预报、控制或模拟等问题的解也是相同的;不难想到,自协方差函数十分相近的随机序列, 它们的预报、控制或模拟等问题的解也是非常接近的. 因此, 不会因为这种识别的"错误"(识别常常是做为近似描述的手段),

而严重影响时序分析的应用效果.

在本章§2中介绍模型初步识别方法的要领时，特别指出对于 z_t 含有非零均值项的情形，需要减掉均值项. 对于 z_t 的均值为常数 $\mu = Ez_t$ 的情形，在实际应用中具有特殊重要性，因此在这里着重考虑这种情况的分析处理方法. 当 μ 不为零而又未知其真值时，从 z_t 中去掉 μ 的方法常常是取

$$w_k = z_k - \hat{\mu} \tag{3.4.4}$$

其中 $\hat{\mu} = \frac{1}{N}\sum_{k=1}^{N} z_k$ 是 z_t 的样本均值. 如果 $\mu = 0$ 时仍用(3.4.4)式来"零化"数据，就会引起估计精度的下降. 可见在对数据 $z_1, z_2, \cdots, z_N$ 进行时序分析时，判断一下 Ez_t 为零与否是很有必要的.

为了判断 $\mu = Ez_t$ 是否为零，我们来分析一下 $\hat{\mu}$ 的性质. 首先

$$E\hat{\mu} = \frac{1}{N}\sum_{k=1}^{N} Ez_k = \frac{1}{N}\sum_{k=1}^{N} \mu = \mu \tag{3.4.5}$$

$$\begin{aligned}
E(\hat{\mu}-\mu)^2 &= E\left[\frac{1}{N}\sum_{k=1}^{N}(z_k-\mu)\right]^2 = \frac{1}{N^2}\sum_{k,j=1}^{N} E(z_k-\mu)(z_j-\mu) \\
&= \frac{1}{N^2}\sum_{k,j=1}^{N} Ew_k w_j = \frac{1}{N^2}\sum_{k,j=1}^{N}\gamma_{k-j} \\
&= \frac{1}{N^2}\sum_{l=1-N}^{N-1}\sum_{k-j=l}\gamma_l = \frac{1}{N^2}\sum_{l=1-N}^{N-1}(N-|l|)\gamma_l \\
&= \frac{1}{N}\sum_{l=1-N}^{N-1}\gamma_l - \frac{1}{N^2}\sum_{l=1-N}^{N-1}|l|\gamma_l
\end{aligned}$$

注意对于 ARMA 序列，$\gamma_k(=\gamma_0\rho_k)$ 总被一负指数函数所控制. 所以

$$\sum_{l=-\infty}^{\infty}|\gamma_l| < \infty, \qquad \sum_{l=-\infty}^{\infty}|l||\gamma_l| < \infty$$

因此再根据前式便得到如下的渐近等式

$$\operatorname{var}(\hat{\mu}) = E(\hat{\mu} - \mu)^2 \approx \frac{1}{N} \sum_{l=-\infty}^{\infty} \gamma_l = \frac{\gamma_0}{N} \sum_{l=-\infty}^{\infty} \rho_l \quad (3.4.6)$$

若 z_t 为正态序列，则 $\hat{\mu}$ 也是正态变量，其均值和渐近方差由 (3.4.5) 和 (3.4.6) 式给出，因此根据正态分布可知

$$P_\mu\{|\hat{\mu} - \mu| \leqslant 3[\operatorname{var}(\hat{\mu})]^{\frac{1}{2}}\} = 99.7\% \quad (3.4.7)$$

若 $\mu = 0$ 上式变为

$$P_0\{|\hat{\mu}| \leqslant 3[\operatorname{var}(\hat{\mu})]^{\frac{1}{2}}\} = 99.7\% \quad (3.4.8)$$

由此便可以判断 $\mu = 0$ 与否．若实际算出的 $\hat{\mu}$ 不满足 $|\hat{\mu}|^2 \leqslant 9\operatorname{var}(\hat{\mu})$，我们有理由否定 $\mu = 0$，而认为 $\mu \neq 0$．事实上如果 μ 真的是零，我们的判断就犯了以假为真的错误，而根据 (3.4.8) 式，犯这种错误的概率只有 0.3%．

下面我们根据 (3.4.6) 式给出几种常用低阶模型的 $\operatorname{var}(\hat{\mu})$ 的公式

AR(1, 0) 模型， $\operatorname{var}(\hat{\mu}) \approx \dfrac{\gamma_0(1 + \rho_1)}{N(1 - \rho_1)} = \dfrac{\gamma_0(1 + \varphi_1)}{N(1 - \varphi_1)}$

AR(2, 0) 模型， $\operatorname{var}(\hat{\mu}) \approx \dfrac{\gamma_0(1 + \rho_1)(1 - 2\rho_1^2 + \rho_2)}{N(1 - \rho_1)(1 - \rho_2)}$

$$= \frac{\gamma_0(1 + \varphi_1 - \varphi_2)(1 + \varphi_2)}{N(1 - \varphi_1 - \varphi_2)(1 - \varphi_2)}$$

MA(0, 1) 模型， $\operatorname{var}(\hat{\mu}) \approx \dfrac{\gamma_0}{N}(1 + 2\rho_1) = \dfrac{\gamma_0}{N} \dfrac{(1 - \theta_1)^2}{1 + \theta_1^2}$

MA(0, 2) 模型 $\operatorname{var}(\hat{\mu}) \approx \dfrac{\gamma_0}{N}(1 + 2\rho_1 + 2\rho_2)$

$$= \frac{\gamma_0(1 - \theta_1 - \theta_2)^2}{N(1 + \theta_1^2 + \theta_2^2)}$$

ARMA(1, 1) 模型 $\operatorname{var}(\hat{\mu}) \approx \dfrac{\gamma_0(\rho_1 - \rho_2 + 2\rho_1^2)}{N(\rho_1 - \rho_2)}$

$$= \frac{\gamma_0(1 + \varphi_1)(1 - \theta_1)^2}{N(1 - \varphi_1)(1 + \theta_1^2 - 2\varphi_1\theta_1)}$$

在求以上诸式时，要用到第二章 §2 中的公式或相应的例子，我们不去一一推导了．当然，在实际应用时，公式中的 ρ_k，φ_k，θ_k 以及

γ_0 一般只能用它的估计值代替. 最后，仍以本节用过的伪随机序列的数值结果作为例子，但是仅考查(2.1.16a)和(2.1.16c)两个模型

(2.1.16c)模型的300个样本数据所计算出的 $\hat{\mu}=0.09$，而

$$\mathrm{var}(\hat{\mu})\approx\frac{\gamma_0}{300}\times\frac{1+0.5}{1-0.5}=\frac{\sigma_a^2}{300[1-(0.5)^2]}\cdot\frac{1.5}{0.5}$$

$$=\frac{4}{100\times 3}=0.013$$

若把 σ_a^2, φ_1 的真实值都当作不知道的量，而以它们的估计值代替时，则

$$\mathrm{var}(\hat{\mu})\approx\frac{1.91}{300}\times\frac{1+0.56}{1-0.56}=0.021$$

以上两式的平方根分别约为0.12和0.14，无论哪一个值都比0.09大，故可以认为 $\mu=0$.

(2.1.16a)模型的300个样本数据所计算出的 $\hat{\mu}=0.001$，而

$$\mathrm{var}(\hat{\mu})\approx\frac{\gamma_0(1-\theta_1)^2}{300(1+\theta_1^2)}=\frac{(1+\theta_1^2)\sigma_a^2(1-\theta_1)^2}{300(1+\theta_1^2)}$$

$$=\frac{(1-0.5)^2}{300}=\frac{1}{1200}=0.00083$$

若将 θ_1 与 σ_a^2 用它们的估计值代替时，则

$$\mathrm{var}(\hat{\mu})\approx\frac{1.23(1-0.38)^2}{300(1+0.38)^2}=0.0011$$

以上两式的平方根分别约为 0.029 和 0.031；无论那一个值都比0.001大，故可以认为 $\mu=0$.

第四章　参数的精估计

本章继续讨论参数估计问题．在模型阶数初步确定以后，上一章介绍的估计模型参数 $\boldsymbol{\varphi}$，$\boldsymbol{\theta}$ 和 σ_a^2 的方法，不要求估计量满足某种最优化约束条件，故可称之为粗估计．本章将要介绍的三种估计，都是要求满足某种最优化约束条件时的估计，故可称为精估计，它们比粗估计的精度高．当样本长度 N 较大时（比如在数百以上），这三种精估计的效果则无甚差别，这时使用最小二乘估计较好，因为这种方法比其它两种方法易懂，计算量又小．当 N 在数百以内时，使用最小平方和估计会有进一步的收益．当 N 很小时（比如在数十以内），采用更复杂的极大似然估计还可能提高一点精度．不过，毕竟由于样本太小，其估计误差还会很大，所以这后一种方法实际上很少使用．本章的三种精估计，不仅在精度上很接近，而且它们的数值解法也十分相近．因此，我们把这些估计的具体解法都放在本章 § 3 中介绍．

§ 1. 最小二乘估计

在这一节里，我们讨论如何把 ARMA 序列的参数估计问题转化为第一章 § 5 中的最小二乘问题．现在分 AR 序列、MA 序列和 ARMA 序列两种不同情况介绍．

一、AR 序列参数的最小二乘估计

将 AR 模型（2.1.8）式改写如下：

$$\begin{cases} w_{p+1} = \varphi_1 w_p + \varphi_2 w_{p-1} + \cdots + \varphi_p w_1 + a_{p+1} \\ w_{p+2} = \varphi_1 w_{p+1} + \varphi_2 w_p + \cdots + \varphi_p w_2 + a_{p+2} \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ w_N = \varphi_1 w_{N-1} + \varphi_2 w_{N-2} + \cdots + \varphi_p w_{N-p} + a_N \end{cases} \tag{4.1.1}$$

如果记 $y_t = w_{p+t}$, $e_t = a_{p+t}$, $-p+1 \leqslant t \leqslant n(n = N-p)$;再令 $\boldsymbol{y}_t = (y_1, y_2, \cdots, y_t)^{\tau}$, $\boldsymbol{\beta} = \boldsymbol{\varphi} = (\varphi_1, \varphi_2, \cdots, \varphi_p)^{\tau}$ 以及

$$f_t(\boldsymbol{\varphi}, \boldsymbol{y}_t) = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p}$$

那么,(4.1.1)式便化成(1.5.4)式的模型了. $\boldsymbol{\varphi}$ 的最小二乘估计 $\hat{\boldsymbol{\varphi}}_L$ 应使下式的 $S(\boldsymbol{\varphi})$ 达到极小,此处

$$S(\boldsymbol{\varphi}) = \sum_{t=1}^{n} \{y_t - f_t(\boldsymbol{\varphi}, \boldsymbol{y}_t)\}^2$$

$$= \sum_{t=1}^{n} (w_{p+t} - \varphi_1 w_{p+t-1} - \varphi_2 w_{p+t-2} - \cdots - \varphi_p w_t)^2$$

为了求上式的极小值,用偏导数方法取

$$\frac{\partial}{\partial \boldsymbol{\varphi}} S(\boldsymbol{\varphi}) \bigg|_{\boldsymbol{\varphi} = \hat{\boldsymbol{\varphi}}_L} = 0$$

于是可得以下方程组

$$\sum_{t=1}^{n} (w_{p+t} - \hat{\varphi}_1^L w_{p+t-1} - \hat{\varphi}_2^L w_{p+t-2} - \cdots - \hat{\varphi}_p^L w_t) w_{p+t-j} = 0$$

$$j = 1, 2, \cdots, p$$

再将上式改写成

$$\frac{\hat{\varphi}_1^L}{N} \sum_{t=1}^{N-p} w_{p+t-1} w_{p+t-j} + \frac{\hat{\varphi}_2^L}{N} \sum_{t=1}^{N-p} w_{p+t-2} w_{p+t-j} + \cdots$$

$$+ \frac{\hat{\varphi}_p^L}{N} \sum_{t=1}^{N-p} w_t w_{p+t-j}$$

$$= \frac{1}{N} \sum_{t=1}^{N-p} w_{p+t} w_{p+t-j}, \quad j = 1, 2, \cdots, p$$

由此线性方程组便可解出 $\boldsymbol{\varphi}$ 的最小二乘估计. 令

$$\hat{\gamma}_{j-i}^L = \frac{1}{N} \sum_{t=1}^{N-p} w_{p+t-i} w_{p+t-j}$$

当 N 较大时上面的方程可近似为

$$\begin{cases} \hat{\varphi}_1^L \hat{\gamma}_0^L + \hat{\varphi}_2^L \hat{\gamma}_1^L + \cdots + \hat{\varphi}_p^L \hat{\gamma}_{p-1}^L = \hat{\gamma}_1^L \\ \hat{\varphi}_1^L \hat{\gamma}_1^L + \hat{\varphi}_2^L \hat{\gamma}_0^L + \cdots + \hat{\varphi}_p^L \hat{\gamma}_{p-2}^L = \hat{\gamma}_2^L \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \hat{\varphi}_1^L \hat{\gamma}_{p-1}^L + \hat{\varphi}_2^L \hat{\gamma}_{p-2}^L + \cdots + \hat{\varphi}_p^L \hat{\gamma}_0^L = \hat{\gamma}_p^L \end{cases} \tag{4.1.2}$$

由此方程求得 $\boldsymbol{\varphi}$ 的最小二乘估计的近似解为

$$\hat{\boldsymbol{\varphi}}_L=\begin{pmatrix}\hat{\varphi}_1^L\\\hat{\varphi}_2^L\\\vdots\\\hat{\varphi}_p^L\end{pmatrix}\cong\begin{pmatrix}\hat{\gamma}_0^L & \hat{\gamma}_1^L & \cdots & \hat{\gamma}_{p-1}^L\\\hat{\gamma}_1^L & \hat{\gamma}_0^L & \cdots & \hat{\gamma}_{p-2}^L\\\cdots & \cdots & \cdots & \cdots\\\hat{\gamma}_{p-1}^L & \hat{\gamma}_{p-2}^L & \cdots & \hat{\gamma}_0^L\end{pmatrix}^{-1}\begin{pmatrix}\hat{\gamma}_1^L\\\hat{\gamma}_2^L\\\vdots\\\hat{\gamma}_p^L\end{pmatrix}\tag{4.1.3}$$

可见在 AR 序列的情形，参数 $\boldsymbol{\varphi}$ 的最小二乘估计 $\hat{\boldsymbol{\varphi}}_L$ 能够通过简单的表达式给出，这是由于 $f_t(\boldsymbol{\varphi},\boldsymbol{y}_t)$ 为 $\boldsymbol{\varphi}$ 的线性函数的缘故．但是，注意 $f_t(\boldsymbol{\varphi},\boldsymbol{y}_t)$ 与量测 $\boldsymbol{y}_t$ 有关，所以 $\hat{\boldsymbol{\varphi}}_L$ 并不是 $\boldsymbol{y}_n$ 的线性函数．

依据

$$a_t=w_t-\varphi_1w_{t-1}-\cdots-\varphi_pw_{t-p}$$

在 $\boldsymbol{\varphi}$ 的估计获得后，a_t 可被估计为

$$\hat{a}_t=w_t-\hat{\varphi}_1^Lw_{t-1}-\cdots-\hat{\varphi}_p^Lw_{t-p}$$

$$t=p+1,p+2,\cdots,N$$

因此，σ_a^2 的最小二乘估计取作

$$\hat{\sigma}_a^2=\frac{1}{N-p}\sum_{t=p+1}^{N}\hat{a}_t^2=\frac{1}{N-p}S(\hat{\boldsymbol{\varphi}}_L)\tag{4.1.4}$$

又当 N 较大时，易见 $\hat{\gamma}_k^L\cong\hat{\gamma}_k$，因此，$\hat{\boldsymbol{\varphi}}_L$ 与 (3.3.1) 式的矩估计 $\hat{\boldsymbol{\varphi}}$ 是十分相近的．

二、MA 与 ARMA 序列参数的最小二乘估计

这两种序列的参数最小二乘估计方法基本相同．我们只叙述 ARMA 序列的估计方法．将 (2.1.5) 式改写成类似 (4.1.1) 式的形式，并注意应用 a_t 的逆转表达式 (2.1.23) 式，可得

$$\begin{aligned}w_{p+t}&=\varphi_1w_{p+t-1}+\varphi_2w_{p+t-2}+\cdots+\varphi_pw_t\\&\quad-\theta_1a_{p+t-1}-\theta_2a_{p+t-2}-\cdots-\theta_qa_{p+t-q}+a_{p+t}\\&=\varphi_1w_{p+t-1}+\varphi_2w_{p+t-2}+\cdots+\varphi_pw_t\\&\quad+\theta_1\sum_{j=0}^{\infty}\pi_jw_{p+t-1-j}+\theta_2\sum_{j=0}^{\infty}\pi_jw_{p+t-2-j}\\&\quad+\cdots+\theta_q\sum_{j=0}^{\infty}\pi_jw_{p+t-q-j}+a_{p+t}\end{aligned}$$

其中记 $\pi_0 = -1$. 再令 $y_t = w_{p+t}$, $e_t = a_{p+t}$, 对于 $t \leqslant n(n = N - p)$,以及 $\boldsymbol{\beta}^\tau = (\boldsymbol{\varphi}^\tau, \boldsymbol{\theta}^\tau)$, $\boldsymbol{y}_t^\tau = (y_1, y_2, \cdots, y_t)$ 和

$$f_t(\boldsymbol{\beta}, \boldsymbol{y}_t) = \sum_{i=1}^{p} \varphi_i y_{t-i} + \sum_{i=1}^{q} \theta_i \sum_{j=0}^{\infty} \pi_j y_{t-i-j}$$

那么,前一式便可化成(1.5.4)式的模型了. 但是,这时 $f_t(\boldsymbol{\beta}, \boldsymbol{y}_t)$ 牵涉到无穷多个 y_k(即 w_k)之值,而我们仅知 $w_1, w_2, \cdots, w_N$ 这 N个数据;为了使用最小二乘估计方法,我们把 $w_k(k \leqslant 0)$ 都取做它们的均值,即 $w_k = Ew_k = 0(k \leqslant 0)$. 于是得到 (1.5.4) 式那样的非线性最小二乘模型(注意 π_j 是 $\boldsymbol{\theta}, \boldsymbol{\varphi}$ 的函数)

$$y_t = f_t(\boldsymbol{\beta}, \boldsymbol{y}_t) + e_t, \quad -p + 1 \leqslant t \leqslant n \tag{4.1.5}$$

其中

$$f_t(\boldsymbol{\beta}, \boldsymbol{y}_t) = \sum_{i=1}^{p} \varphi_i y_{t-i} + \sum_{i=1}^{q} \theta_i \sum_{j=0}^{p+t-i-1} \pi_j y_{t-i-j} \tag{4.1.6}$$

由此确定的最小二乘估计 $\hat{\boldsymbol{\beta}}_L$, 实际上相当于根据无穷长度样本序列 $w_N, w_{N-1}, w_{N-2}, \cdots, w_1, 0, 0, \cdots$, 对 $\boldsymbol{\beta}$ 所做的最小二乘估计. $w_0, w_{-1}, \cdots$ 这些量的样本值并不知道,只是为了使得方法便于进行计算,才用零值代替它们. 即使如此,模型(4.1.5)式的最小二乘估计也无法简单求解,只能用数值解法求解. 为此我们令 $\boldsymbol{w}_k^\tau = (w_1, w_2, \cdots, w_k)$ 依(2.1.23)我们取 a_k 的近似估计如下

$$\hat{a}_k = \hat{a}(\boldsymbol{\beta}, \boldsymbol{w}_k) = -\sum_{j=0}^{k-1} \pi_j w_{k-j} = w_k - \sum_{j=1}^{k-1} \pi_j w_{k-j}, \quad k \geqslant 1$$

$$\hat{a}_k = 0, \quad k \leqslant 0 \tag{4.1.7}$$

易见当 $k \geqslant p + 1$ 时

$$\hat{a}_k = w_k - f_{k-p}(\boldsymbol{\beta}, \boldsymbol{y}_{k-p}) = y_{k-p} - f_{k-p}(\boldsymbol{\beta}, \boldsymbol{y}_{k-p})$$

于是,与 $S(\boldsymbol{\varphi})$ 相似地定义残差平方和为

$$S(\boldsymbol{\beta}) = \sum_{k=p+1}^{N} \hat{a}_k^2 \tag{4.1.8}$$

在必要时,我们用记号 $\hat{a}_k(\boldsymbol{\beta}, \boldsymbol{w}_k)$, 以标明估计 a_k 时所用的参数为 $\boldsymbol{\beta}$, 所用的数据为 $\boldsymbol{w}_k$. 所谓 $\boldsymbol{\beta}$ 的最小二乘估计,就是寻求这样

的参数 $\hat{\boldsymbol{\beta}}_L$，以使 $\hat{a}_k(\hat{\boldsymbol{\beta}}_L, \boldsymbol{w}_k)$ 的平方和 $S(\hat{\boldsymbol{\beta}}_L)$ 达到 $S(\boldsymbol{\beta})$ 的极小值．为了具体求解 $\hat{\boldsymbol{\beta}}_L$，一般并不用(4.1.7)式逐个计算 $\hat{a}_k$，这样计算量太大，而采用递推算法，即用 ARMA 序列的递推公式

$$\hat{a}_k = w_k - \sum_{i=1}^{p} \varphi_i w_{k-i} + \sum_{i=1}^{q} \theta_i \hat{a}_{k-i}, \quad k = 1, 2, \cdots, N \qquad (4.1.9)$$

其中 $\hat{a}_k = 0(k \leqslant 0)$．这样，对每一具体参数值 $\boldsymbol{\beta}$，便可由(4.1.9)式与(4.1.8)式算出 $\hat{a}_k(k = 1, 2, \cdots, N)$和 $S(\boldsymbol{\beta})$．在所有 $S(\boldsymbol{\beta})$ 中选出极小的一个，相应的 $\hat{\boldsymbol{\beta}}_L$ 即为最小二乘估计．

求(4.1.8)式的极小值之解是一个普通的求极值问题，读者可以在动态规划中找到各种各样的数值解法．我们将针对这里的具体问题提供几种特殊的求解方法，这些方法将在 § 3 中介绍．最后，我们举一个最简单的例子，以说明(4.1.8)式极小值求解的复杂性．

例如求 MA(0, 1) 序列的参数的最小二乘估计．

$$\hat{a}_0 = 0$$

$$\hat{a}_1 = w_1 + \theta_1 \hat{a}_0 = w_1$$

$$\hat{a}_2 = w_2 + \theta_1 \hat{a}_1 = w_2 + \theta_1 w_1$$

$$\cdots\cdots\cdots\cdots$$

$$\hat{a}_k = w_k + \theta_1 \hat{a}_{k-1} = w_k + \theta_1 w_{k-1} + \theta_1^2 w_{k-2} + \cdots + \theta_1^{k-1} w_1 = \sum_{j=0}^{k-1} \theta_1^j \omega_{k-j}$$

于是

$$S(\theta_1) = \sum_{k=1}^{N} \hat{a}_k^2 = \sum_{k=1}^{N} \left(\sum_{j=0}^{k-1} \theta_1^j w_{k-j} \right)^2$$

由此可见，即使对于 MA(0, 1) 序列而言，求出使 $S(\theta_1)$ 达极小值的 $\hat{\theta}_L$ 来也不是很简单的事情．

§ 2. 最小平方和估计

一、最小平方和估计的概念

在前一节中考虑 MA 或 ARMA 序列参数的最小二乘估计

时，我们为了使其方法便于实际计算，拟造了 $k \leq 0$ 各时刻的 w_k 样本值，即代之以 $w_k = 0(k \leq 0)$. 在(4.1.7)式定义 a_k 的估计时亦如此. 这样人为地添加了假数据 $w_k = 0(k \leq 0)$，无疑会影响估计的精度. 为了只根据样本值数据 $w_1, w_2, \cdots, w_N$，对参数做出另一种精估计，我们首先需要知道，在有穷样本值和给定的参数值下，如何对 a_k 做出最小方差估计(见第一章 § 5)，即

$$\hat{a}_k = \hat{a}_k(\boldsymbol{\beta}, \boldsymbol{w}_N) = E_{\boldsymbol{\beta}}(a_k | w_1, w_2, \cdots, w_N) \qquad (4.2.1)$$

其中 $\boldsymbol{w}_N^\tau = (w_1, w_2, \cdots, w_N)$. 显而易见，$\hat{a}_k$ 是 $\boldsymbol{\beta}$ 和 $\boldsymbol{w}_N$ 的函数. 经初步识别后已判断 w_t 是 ARMA 序列，所以，在给出一组参数值 $\boldsymbol{\beta}$ 后，按照第二章 § 2 所述，这就等于给出了序列 w_t 和序列 a_t 的概率分布. 根据第一章 § 5 中关于最小方差估计的各条性质，以及第二章 § 2 的有关公式，便可以确定出(4.2.1)式的 $\hat{a}_k$. 在本节二中将详细给出计算 $\hat{a}_k$ 的公式. 按照上一节类似的办法，定义如下的函数

$$S(\boldsymbol{\beta}) = \sum_{k=-\infty}^{N} \hat{a}_k^2 = \sum_{k=-\infty}^{N} \{E_{\boldsymbol{\beta}}(a_k | \boldsymbol{w}_N)\}^2 \qquad (4.2.2)$$

称 $S(\boldsymbol{\beta})$ 为 ARMA 序列的平方和函数. 对于固定的样本值 $\boldsymbol{w}_N$ 而言，$S(\boldsymbol{\beta})$ 是 $\boldsymbol{\beta}$ 的函数. 使 $S(\boldsymbol{\beta})$ 达到极小值的 $\hat{\boldsymbol{\beta}}$ 就称为参数 $\boldsymbol{\beta}$ 的最小平方和估计. 在本章 § 4 中将指出，这是一种近似的极大似然估计. 相应的 σ_a^2 的最小平方和估计为

$$\hat{\sigma}_a^2 = \frac{S(\hat{\boldsymbol{\beta}})}{N}$$

下面，仍旧分 AR 序列、MA 和 ARMA 序列两种情况进行论讨.

1. AR 序列参数的最小平方和估计.

根据 ARMA 序列的性质，将(4.2.2)式的 $S(\boldsymbol{\beta})$ 改写成以下的形式(证明见附录 § 3):

$$S(\boldsymbol{\beta}) = \sigma_a^2 \boldsymbol{w}_N^\tau \Gamma_N^{-1} \boldsymbol{w}_N = \boldsymbol{w}_N^\tau M_N \boldsymbol{w}_N \qquad (4.2.3)$$

其中 $\boldsymbol{w}_N$ 仍表示数据向量 $(w_1, w_2, \cdots, w_N)^\tau$，$\Gamma_N$ 为 $\boldsymbol{w}_N$ 的协方差阵(见(1.3.9)式)，以及

$$M_N = \sigma_a^2 \Gamma_N^{-1} = \sigma_a^2 (E_{\boldsymbol{\beta}} \boldsymbol{w}_N \boldsymbol{w}_N^\tau)^{-1} \qquad (4.2.4)$$

如果 w_t 是 AR 序列则 $\boldsymbol{\beta}=\boldsymbol{\varphi}$，根据附录 § 3 的推导，可知

$$S(\boldsymbol{\varphi})=\boldsymbol{w}_N^\tau M_N \boldsymbol{w}_N=\sum_{ij=0}^{p}\varphi_i\varphi_j d_{ij} \tag{4.2.5}$$

在这里我们采用了记号 $\varphi_0\equiv -1$．当 $N\geqslant 2p$ 时，

$$d_{ij}=\sum_{k=1+j}^{N-i} w_{k+i-j}w_k=d_{ji},\qquad i,j=0,1,\cdots,p \tag{4.2.6}$$

当 $N<2p$ 时，d_{ij} 的表达式可见附录 § 3．在实际应用中，N 远远大于 $2p$，所以这里只给出 $N\geqslant 2p$ 时 d_{ij} 的表达式． 我们用矩阵形式表达 (4.2.5) 式，为此令 p 阶方阵

$$D=\begin{pmatrix} d_{11} & d_{12} & \cdots & d_{1p}\\ d_{21} & d_{22} & \cdots & d_{2p}\\ \cdots & \cdots & \cdots & \cdots\\ d_{p1} & d_{p2} & \cdots & d_{pp}\end{pmatrix}$$

向量 $\boldsymbol{d}=(d_{01},d_{02},\cdots,d_{0p})^\tau$，那么 (4.2.5) 式又可写做

$$S(\boldsymbol{\varphi})=\boldsymbol{\varphi}^\tau D\boldsymbol{\varphi}-2\boldsymbol{\varphi}^\tau\boldsymbol{d}=d_{00} \tag{4.2.7}$$

此式为 $\boldsymbol{\varphi}$ 的二次型，使其达到极小值的解为

$$\hat{\boldsymbol{\varphi}}_S=D^{-1}\boldsymbol{d} \tag{4.2.8}$$

σ_a^2 的最小平方和估计定义为

$$\hat{\sigma}_a^2=\frac{1}{N}S(\hat{\boldsymbol{\varphi}}_S)=\frac{1}{N}(\hat{\boldsymbol{\varphi}}_S^\tau D\hat{\boldsymbol{\varphi}}_S-2\hat{\boldsymbol{\varphi}}_S^\tau\boldsymbol{d}+d_{00})$$
$$=\frac{1}{N}(d_{00}-\boldsymbol{d}^\tau D^{-1}\boldsymbol{d}) \tag{4.2.9}$$

(4.2.8) 式和 (4.2.9) 式就是 AR(p, 0) 模型参数的最小平方和估计的表达式． 与 AR 模型参数的最小二乘估计相比较，它们有同样简单的表达形式．而且当 N 较大时，

$$\frac{1}{N}d_{ij}=\frac{1}{N}\sum_{k=1+j}^{N-i}w_{k+i-j}w_k\approx\frac{1}{N}\sum_{t=1}^{N-p}w_{p+t-i}w_{p+t-j}=\hat{\gamma}_{j-i}^{\tau}$$

所以 (4.2.8) 式与 (4.1.3) 式也是十分相近的，从而 (4.2.9) 式与 (4.1.4) 式也很相近． 同样，在 N 适当大以后，这两种估计与第三章 § 3 的矩估计也是非常接近的．由此可见，AR 模型的参数估计

是容易获得的；若N比较大，其各种估计的结果都相差不多．而MA 或 ARMA 模型则不然，它们的参数精估计都是十分烦琐的．因此，有人认为在描述真实的随机序列时，应当尽量采用AR 模型；即使真实序列是 MA 或 ARMA 序列，也可以用高阶自回归模型近似描述．

2. MA 与 ARMA 序列参数的最小平方和估计．

对于 MA 序列而言，由于 $a_{-q-j}(j \geqslant 0)$与 $w_1, w_2, \cdots, w_N$ 独立，因此(4.2.1)式的 $\hat{a}_{-q-j} = 0(j \geqslant 0)$．于是(4.2)式为

$$S(\boldsymbol{\beta}) = S(\boldsymbol{\theta}) = \sum_{k=-q+1}^{N} \hat{a}_k^2 \tag{4.2.10}$$

虽然上式的求和项只有有穷项，但是，欲求 $S(\boldsymbol{\theta})$ 的极小值仍须采用数值解法．这是因为，首先每个 $\hat{a}_k$ 都是 $\boldsymbol{\theta}$ 的复杂函数形式，给出 $\boldsymbol{\theta}$ 后，要用数值计算方法才能求出相应的 $\hat{a}_k = \hat{a}_k(\boldsymbol{\theta}, \boldsymbol{w}_N)$ 和 $S(\boldsymbol{\theta})$ 之值．其次，要在所有可能的 $\boldsymbol{\theta}$ 值中选出使 $S(\boldsymbol{\theta})$ 达极小的解，也只能用数值计算方法求近似值．关于 $\hat{a}_k$ 或 $S(\boldsymbol{\theta})$ 的计算方法在下一小节专门讨论．至于求 $S(\boldsymbol{\theta})$ 的极小值的解法，在本章§3 中再介绍．

对于 ARMA 序列而言，其步骤与 MA 序列相同，只是在计算平方和函数 $S(\boldsymbol{\beta})$ 时，须要用有穷和近似无穷和，即

$$S(\boldsymbol{\beta}) = \sum_{k=-\infty}^{N} \hat{a}_k^2 \cong \sum_{k=-Q}^{N} \hat{a}_k^2 \tag{4.2.11}$$

其中 Q 是依具体参数 $\boldsymbol{\beta}$ 而定的正整数，使得 $\hat{a}_k^2(k < -Q)$ 为可忽略量．

二、平方和函数的计算方法

为了给出上一小节所定义的 $\hat{a}_k$ 和 $S(\boldsymbol{\beta})$ 的数值计算方法，我们先来讨论 ARMA 序列的一种对称性质．设 x_t 为 ARMA 序列，即满足差分方程

$$\varphi(B)x_t = \theta(B)a_t \tag{4.2.12}$$

若N是一任意给定的整数，令 $y_t = x_{N+1-t}$，则 y_t 与 x_t 的自协方

差函数完全相同．因此 y_t 也是 ARMA 序列，而且与 x_t 有相同的阶数和相同的参数．为了区别于 x_t，我们把 y_t 的模型写做

$$\varphi(B)y_t = \theta(B)b_t \tag{4.2.13}$$

其中 b_t 也是白噪声序列，而且 $Eb_t^2 = Ea_t^2$．我们称 y_t 为 x_t 的一个对称序列．这说明在研究平稳序列的相关结构时，时间自变量 t 的正序与反序没有差别．设已给数据 $\boldsymbol{w}_N = (w_1, w_2, \cdots, w_N)^\tau$，经模型初步识别后被定为 ARMA$(p, q)$ 序列，更确切地说，数据列 $w_1, w_2, \cdots, w_N$ 被认为是满足某一(4.2.12)式的 x_t 的一段样本值．根据上述的对称性，数据列 $w_N, w_{N-1}, \cdots, w_1$ 就应是相应的(4.2.13)式的 y_t 的一段样本值．即 $w_N, w_{N-1}, \cdots, w_1$ 与 $w_1, w_2, \cdots, w_N$ 以相同概率做为某一 ARMA 序列的一段样本出现．

现在引进记号 $\boldsymbol{w}_N^* = (w_N, w_{N-1}, \cdots, w_1)$．假定根据初步识别结果，已确定数据 $w_1, w_2, \cdots, w_N$ 是 x_t 的一段样本，用 $\boldsymbol{\beta}$ 表示 x_t 的参数向量．于是，用样本值 $\boldsymbol{w}_N$ 对 x_t 做的最小方差估计可以表示成

$$\hat{x}_t = E_\beta(x_t | w_1, w_2, \cdots, w_N) = E_\beta(x_t | \boldsymbol{w}_N)$$

类似地有

$$\hat{a}_t = E_\beta(a_t | w_1, w_2, \cdots, w_N) = E_\beta(a_t | \boldsymbol{w}_N)$$

当把 $\boldsymbol{w}_N^*$ 用于 x_t 的对称序列 y_t 时，类似地又有

$$\hat{y}_t = E_\beta(y_t | w_N, w_{N-1}, \cdots, w_1) = E_\beta(y_t | \boldsymbol{w}_N^*)$$

$$\hat{b}_t = E_\beta(b_t | w_N, w_{N-1}, \cdots, w_1) = E_\beta(b_t | \boldsymbol{w}_N^*)$$

这些量之间有一定的关系．根据第一章 §5 所介绍的最小方差估计的性质，以及 x_t, y_t 各自的模型和自协方差函数相同这一事实，可知

$$\varphi(B)\hat{x}_t = \theta(B)\hat{a}_t$$

$$\varphi(B)\hat{y}_t = \theta(B)\hat{b}_t$$

即

$$\hat{a}_t = \hat{x}_t - \varphi_1\hat{x}_{t-1} - \cdots - \varphi_p\hat{x}_{t-p} + \theta_1\hat{a}_{t-1} + \cdots + \theta_q\hat{a}_{t-q} \tag{4.2.14}$$

$$\hat{b}_t = \hat{y}_t - \varphi_1\hat{y}_{t-1} - \cdots - \varphi_p\hat{y}_{t-p} + \theta_1\hat{b}_{t-1} + \cdots + \theta_q\hat{b}_{t-q} \tag{4.2.15}$$

又由最小方差估计的性质可知

$$\hat{x}_k = w_k, \quad \hat{y}_k = w_{N-k+1}, \quad k = 1,2,\cdots,N \tag{4.2.16}$$

$$\hat{a}_{N+k} = \hat{b}_{N+k} = 0, \quad k = 1,2,\cdots \tag{4.2.17}$$

令 $\boldsymbol{x}_N^\tau = (x_1, x_2, \cdots, x_N)$，$\boldsymbol{y}_N^\tau = (y_1, y_2, \cdots, y_N)$；方阵 T 如第二章 § 2 所定义（见 (2.2.23) 式后边的 T）. 由对称性易见

$$T(E\boldsymbol{y}_N\boldsymbol{y}_N^\tau)T = E\boldsymbol{y}_N\boldsymbol{y}_N^\tau = E\boldsymbol{x}_N\boldsymbol{x}_N^\tau = T(E\boldsymbol{x}_N\boldsymbol{x}_N^\tau)T$$

$$\boldsymbol{w}_N^* = T\boldsymbol{w}_N$$

$$\begin{aligned}(Ex_k\boldsymbol{x}_N^\tau)T &= (Ey_k\boldsymbol{y}_N^\tau)T = (Ey_k\boldsymbol{y}_N^\tau T)\\ &= (\gamma_{k-N}, \gamma_{k-N+1}, \cdots, \gamma_{k-1})\\ &= (\gamma_{N-k}, \gamma_{N-k-1}, \cdots, \gamma_{1-k})\\ &= Ey_{N-k+1}\boldsymbol{y}_N^\tau = Ex_{N-k+1}\boldsymbol{x}_N^\tau\end{aligned}$$

从而又有

$$\begin{aligned}\hat{x}_k &= (Ex_k\boldsymbol{x}_N^\tau)(E\boldsymbol{x}_N\boldsymbol{x}_N^\tau)^{-1}\boldsymbol{w}_N = (Ey_k\boldsymbol{y}_N^\tau)(E\boldsymbol{y}_N\boldsymbol{y}_N^\tau)^{-1}\boldsymbol{w}_N\\ &= (Ey_{N-k+1}\boldsymbol{y}_N^\tau)(E\boldsymbol{y}_N\boldsymbol{y}_N^\tau)^{-1}\boldsymbol{w}_N^* = \hat{y}_{N-k+1}\end{aligned} \tag{4.2.18}$$

如果 x_t 是 MA 模型之解时，则还有

$$\hat{x}_{-k} = \hat{y}_{-k} = \hat{a}_{-k} = \hat{b}_{-k} = 0, \quad k = q, q+1, \cdots \tag{4.2.19}$$

根据这些公式，便可以建立 $\hat{a}_t$ 逐步迭代的数值计算方法. 下面分别就 MA 与 ARMA 两种模型仔细加以讨论.

1. MA 模型的 $\hat{a}_t$ 的迭代计算公式.

第一步的迭代初值可取 $\hat{b}_{-q+1}^{(1)} = \hat{b}_{-q+2}^{(1)} = \cdots = \hat{b}_0^{(1)} = 0$，根据公式 (4.2.15)，(4.2.16)，对于 $k = 1, 2, \cdots, N$，$\hat{b}_k^{(i)}$ 可以从 $\hat{b}_{-q+1}^{(i)}, \hat{b}_{-q+2}^{(i)}, \cdots, \hat{b}_0^{(i)}$ 迭代计算，即

$$\hat{b}_k^{(i)} = w_{N-k+1} + \theta_1\hat{b}_{k-1}^{(i)} + \cdots + \theta_q\hat{b}_{k-q}^{(i)}, \quad k = 1, 2, \cdots, N \tag{4.2.20}$$

再利用 (4.2.15)，(4.2.18) 和 (4.2.17) 式建立以下公式：

$$\begin{cases}\hat{x}_0^{(i)} = \hat{y}_{N+1}^{(i)} = -\theta_1\hat{b}_N^{(i)} - \theta_2\hat{b}_{N-1}^{(i)} - \cdots - \theta_q\hat{b}_{N-q+1}^{(i)}\\ \hat{x}_{-1}^{(i)} = \hat{y}_{N+2}^{(i)} = -\theta_2\hat{b}_N^{(i)} - \theta_3\hat{b}_{N-1}^{(i)} - \cdots - \theta_q\hat{b}_{N-q+2}^{(i)}\\ \cdots\cdots\cdots\cdots\cdots\cdots\\ \hat{x}_{-q+1}^{(i)} = \hat{y}_{N+q}^{(i)} = -\theta_q\hat{b}_N^{(i)}\end{cases} \tag{4.2.21}$$

用这些式子可以由 $\hat{b}_1^{(i)}, \hat{b}_2^{(i)}, \cdots, \hat{b}_N^{(i)}$ 算出 $\hat{y}_{N+1}^{(i)}, \hat{y}_{N+2}^{(i)}, \cdots, \hat{y}_{N+q}^{(i)}$ 和 $\hat{x}_0^{(i)}, \hat{x}_{-1}^{(i)}, \hat{x}_{-2}^{(i)}, \cdots, \hat{x}_{-q+1}^{(i)}$. 再根据 (4.2.14) 和 (4.2.19) 式，建立如下递推式：

$$\begin{cases} \hat{a}_{-q+1}^{(i)} = \hat{x}_{-q+1}^{(i)} \\ \hat{a}_{-q+2}^{(i)} = \hat{x}_{-q+2}^{(i)} + \theta_1 \hat{a}_{-q+1}^{(i)} \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ \hat{a}_0^{(i)} = \hat{x}_0^{(i)} + \theta_1 \hat{a}_{-1}^{(i)} + \cdots + \theta_{q-1} \hat{a}_{-q+1}^{(i)} \end{cases} \tag{4.2.22}$$

用这些式子可以由 $\hat{x}_{-q+1}^{(i)}, \hat{x}_{-q+2}^{(i)}, \cdots, \hat{x}_0^{(i)}$ 迭代算出 $\hat{a}_{-q+1}^{(i)}, \hat{a}_{-q+2}^{(i)}, \cdots, \hat{a}_0^{(i)}$. 类似地，与 (4.2.20) 式对称的有

$$\hat{a}_k^{(i)} = w_k + \theta_1 \hat{a}_{k-1}^{(i)} + \cdots + \theta_q \hat{a}_{k-q}^{(i)}, \qquad k = 1, 2, \cdots, N \tag{4.2.23}$$

与 (4.2.21) 式对称的有

$$\begin{cases} \hat{y}_0^{(i+1)} = \hat{x}_{N+1}^{(i)} = -\theta_1 \hat{a}_N^{(i)} - \theta_2 \hat{a}_{N-1}^{(i)} - \cdots - \theta_q \hat{a}_{N-q+1}^{(i)} \\ \hat{y}_{-1}^{(i+1)} = \hat{x}_{N+2}^{(i)} = -\theta_2 \hat{a}_N^{(i)} - \theta_3 \hat{a}_{N-1}^{(i)} - \cdots - \theta_q \hat{a}_{N-q+2}^{(i)} \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \hat{y}_{-q+1}^{(i+1)} = \hat{x}_{N+q}^{(i)} = -\theta_q \hat{a}_N^{(i)} \end{cases} \tag{4.2.24}$$

最后，由与 (4.2.22) 式对称的公式

$$\begin{cases} \hat{b}_{-q+1}^{(i+1)} = \hat{y}_{-q+1}^{(i+1)} \\ \hat{b}_{-q+2}^{(i+1)} = \hat{y}_{-q+2}^{(i+1)} + \theta_1 \hat{b}_{-q+1}^{(i+1)} \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \hat{b}_0^{(i+1)} = \hat{y}_0^{(i+1)} + \theta_1 \hat{b}_{-1}^{(i+1)} + \cdots + \theta_{q-1} \hat{b}_{-q+1}^{(i+1)} \end{cases} \tag{4.2.25}$$

可以算出下步的迭代初值 $\hat{b}_{-q+1}^{(i+1)}, \hat{b}_{-q+2}^{(i+1)}, \cdots, \hat{b}_0^{(i+1)}$. 重复以上诸式，计算次序如下：(4.2.20) → (4.2.21) → (4.2.22) → (4.2.23) → (4.2.24) → (4.2.25)，又可得到更新的 $\hat{b}_{-q+1}^{(i+2)}, \hat{b}_{-q+2}^{(i+2)}, \cdots, \hat{b}_0^{(i+2)}$. 这样反复迭代下去，直到 $|\hat{a}_k^{(j)} - \hat{a}_k^{(j-1)}|(-q < -k \leqslant N)$ 可以被忽略时，便停止迭代，并以 $\hat{a}_k^{(j)}(-q < k \leqslant N)$ 作为 $\hat{a}_k$ 的近似值. 这时 $S(\boldsymbol{\theta})$ 的近似值为

$$S(\boldsymbol{\theta}) \cong \sum_{k=-q+1}^{N} (\hat{a}_k^{(j)})^2$$

2. ARMA 模型的 $\hat{a}_k$ 的迭代计算公式.

建立 ARMA 模型的 $\hat{a}_k$ 的迭代公式与 MA 模型的情形完全类似,只是当 $k \leqslant 0$ 时不仅考虑 $k = 0, -1, \cdots, -q+1$, 而且要考虑到 $k = -q, -q-1, \cdots, -Q$, 使得在 $k < -Q$ 以后, $\hat{x}_k$, $\hat{y}_k$, $\hat{b}_k$ 和 $\hat{a}_k$ 等可以被忽略. 这里的 Q 并不是一个固定的正整数,而是在实际计算时适当掌握的数.

第一步的迭代初值一律取为零, 即 $\hat{b}_k^{(1)} = \hat{y}_k^{(1)} = 0\ (k \leqslant 0)$. 由 $\hat{y}_{-Q}^{(i)}, \hat{y}_{-Q+1}^{(i)}, \cdots, \hat{y}_{-Q+p-1}^{(i)}$ 和 $\hat{b}_{-Q}^{(i)}, \hat{b}_{-Q+1}^{(i)}, \cdots, \hat{b}_{-Q+q}^{(i)}$ 迭代计算 $\hat{b}_k^{(i)}(k \geqslant -Q)$ 时,应用与 (4.2.20) 式相似的迭代公式:

$$\begin{aligned}\hat{b}_k^{(i)} = \hat{y}_k^{(i)} - \varphi_1 \hat{y}_{k-1}^{(i)} - \cdots - \varphi_p \hat{y}_{k-p}^{(i)} + \theta_1 \hat{b}_{k-1}^{(i)} \\ + \cdots + \theta_q \hat{b}_{k-q}^{(i)}, \qquad -Q \leqslant k \leqslant N \end{aligned} \quad (4.2.26)$$

其中 $\hat{y}_k^{(i)} = w_{N-k+1}(1 \leqslant k \leqslant N)$, $\hat{b}_k^{(i)} = \hat{y}_k^{(i)} = 0(k < -Q)$. 与 (4.2.21) 式类似的公式为

$$\begin{cases}\hat{x}_0^{(i)} = \hat{y}_{N+1}^{(i)} = \varphi_1 \hat{y}_N^{(i)} + \varphi_2 \hat{y}_{N-1}^{(i)} + \cdots + \varphi_p \hat{y}_{N-p+1}^{(i)} \\ \qquad\qquad - \theta_1 \hat{b}_N^{(i)} - \cdots - \theta_q \hat{b}_{N-q+1}^{(i)} \\ \hat{x}_{-1}^{(i)} = \hat{y}_{N+2}^{(i)} = \varphi_1 \hat{y}_{N+1}^{(i)} + \varphi_2 \hat{y}_N^{(i)} + \cdots + \varphi_p \hat{y}_{N-p+2}^{(i)} \\ \qquad\qquad - \theta_2 \hat{b}_N^{(i)} - \cdots - \theta_q \hat{b}_{N-q+2}^{(i)} \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \hat{x}_{-q+1}^{(i)} = \hat{y}_{N+q}^{(i)} = \varphi_1 \hat{y}_{N+q-1}^{(i)} + \varphi_2 \hat{y}_{N+q-2}^{(i)} + \cdots \\ \qquad\qquad + \varphi_p \hat{y}_{N+q-p}^{(i)} - \theta_q \hat{b}_N^{(i)} \\ \hat{x}_{-k+1}^{(i)} = \hat{y}_{N+k}^{(i)} = \varphi_1 \hat{y}_{N+k-1}^{(i)} + \varphi_2 \hat{y}_{N+k-2}^{(i)} + \cdots \\ \qquad\qquad + \varphi_p \hat{y}_{N+k-p}^{(i)}, \quad Q+1 \geqslant k > q \end{cases} \quad (4.2.27)$$

与 (4.2.22), (4.2.23) 式类似的公式合并为

$$\begin{aligned}\hat{a}_k^{(i)} = \hat{x}_k^{(i)} - \varphi_1 \hat{x}_{k-1}^{(i)} - \cdots - \varphi_p \hat{x}_{k-p}^{(i)} + \theta_1 \hat{a}_{k-1}^{(i)} + \cdots \\ + \theta_q \hat{a}_{k-q}^{(i)}, \qquad -Q \leqslant k \leqslant N \end{aligned} \quad (4.2.28)$$

这里须注意, 当 $k < -Q$ 时, $\hat{x}_k^{(i)} = \hat{a}_k^{(i)} = 0$, 即认为它们是可忽略的. 又依 (4.2.16) 式, $\hat{x}_k^{(i)} = w_k\ (1 \leqslant k \leqslant N, i \geqslant 1)$. 与 (4.2.24) 式 相应的公式可写做

$$
\begin{cases}
\hat{y}_0^{(i+1)} = \hat{z}_{N+1}^{(i)} = \varphi_1 \hat{z}_N^{(i)} + \cdots + \varphi_p \hat{z}_{N-p+1}^{(i)} - \theta_1 \hat{a}_N^{(i)} \\
\qquad\qquad - \cdots - \theta_q \hat{a}_{N-q+1}^{(i)} \\
\hat{y}_{-1}^{(i+1)} = \hat{z}_{N+2}^{(i)} = \varphi_1 \hat{z}_{N+1}^{(i)} + \cdots + \varphi_p \hat{z}_{N-p+2}^{(i)} - \theta_2 \hat{a}_N^{(i)} \\
\qquad\qquad - \cdots - \theta_q \hat{a}_{N-q+2}^{(i)} \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
\hat{y}_{-q+1}^{(i+1)} = \hat{z}_{N+q}^{(i)} = \varphi_1 \hat{z}_{N+q-1}^{(i)} + \cdots + \varphi_p \hat{z}_{N+q-p}^{(i)} \\
\qquad\qquad - \theta_q \hat{a}_N^{(i)} \\
\hat{y}_{-k+1}^{(i+1)} = \hat{z}_{N+k}^{(i)} = \varphi_1 \hat{z}_{N+k-1} + \cdots + \varphi_p \hat{z}_{N+k-p}^{(i)} \\
\qquad\qquad Q + 1 \geqslant k > q
\end{cases}
\tag{4.2.29}
$$

最后，与(4.2.25)式相应的有

$$
\hat{b}_k^{(i+1)} = \hat{y}_k^{(i+1)} - \varphi_1 \hat{y}_{k-1}^{(i+1)} - \cdots - \varphi_p \hat{y}_{k-p}^{(i+1)} + \theta_1 \hat{b}_{k-1}^{(i+1)} + \cdots + \theta_q \hat{b}_{k-q}^{(i+1)}, \quad -Q \leqslant k \leqslant N \tag{4.2.30}
$$

其中 $\hat{y}_k^{(i+1)} = \hat{b}_k^{(i+1)} = 0 \ (k < -Q)$ 和 $\hat{y}_k^{(i+1)} = w_{N-k+1} \ (1 \leqslant k \leqslant N)$. 这时(4.2.30)式与(4.2.26)式是相同的. 从(4.2.26)式到(4.2.30)式就构成了 $\hat{b}_k^{(i+1)}$ 的完整的循环迭代计算公式. 到某一轮迭代值使 $|\hat{a}_k^{(i)} - \hat{a}_k^{(i-1)}| \ (1 \leqslant k \leqslant N)$ 都可以被忽略时，便停止迭代，并以 $\hat{a}_k^{(i)}$ 做为 $\hat{a}_k$ 的近似值. 这时 ARMA 模型的平方和函数近似值为

$$
S(\boldsymbol{\beta}) \cong \sum_{k=-Q}^{N} (\hat{a}_k^{(i)})^2
$$

为了熟悉上述的方法，我们举一个十分简单的数值例子. 以 ARMA(1, 1) 模型为例，并且只对一组参数值 $\boldsymbol{\beta}^{\tau} = (0.3, 0.6)$ 算出平方和函数 $S(\boldsymbol{\beta})$ 的值，所取的数据为 $\boldsymbol{w}_N = \boldsymbol{w}_4 = (-0.5, 0, 0.5, 0)^{\tau}$. 我们将在表 4.2.1 中给出计算步骤和数值结果. 表中箭头所示的方向表示计算的先后次序；箭头起端的阿拉伯号码代表在本段中数值计算所用的公式，它们分别是：

① → 使用公式 $\hat{b}_k^{(i)} = \hat{y}_k^{(i)} - 0.3\hat{y}_{k-1}^{(i)} + 0.6\hat{b}_{k-1}^{(i)}$

$$-2 \leqslant k \leqslant 4$$

② 使用公式 $\hat{y}_{N+1}^{(i)} = 0.3\hat{y}_N^{(i)} - 0.6\hat{b}_N^{(i)}$

③ → 使用公式 $\hat{y}_{N+k}^{(i)} = 0.3\hat{y}_{N+k-1}^{(i)}, \qquad 2 \leqslant k \leqslant 3$

④ → 使用公式 $\hat{x}_{-k}^{(i)} = \hat{y}_{N+k+1}^{(i)}$,　　$0 \leqslant k \leqslant 2$

⑤ → 使用公式 $\hat{a}_k^{(i)} = \hat{x}_k^{(i)} - 0.3\hat{x}_{k-1}^{(i)} + 0.6\hat{a}_{k-1}^{(i)}$

$$-2 \leqslant k \leqslant 4$$

⑥　使用公式 $\hat{x}_{N+1}^{(i)} = 0.3\hat{x}_N^{(i)} - 0.6\hat{a}_N^{(i)}$

⑦ → 使用公式 $\hat{x}_{N+k}^{(i)} = 0.3\hat{x}_{N+k-1}^{(i)}$,　　$2 \leqslant k \leqslant 3$

⑧ → 使用公式 $\hat{y}_{-k}^{(i+1)} = \hat{x}_{N+k+1}^{(i)}$,　　$0 \leqslant k \leqslant 2$

此外，在表一中把数据值、零近似值用括号“( )”括起来，区别出它们是已知量或近似已知量. 在明显只取零值的地方，也不列入计算线，比如 $\hat{a}_k^{(i)} = \hat{b}_k^{(i)} = 0\ (k > N)$. 详细结果见表 4.2.1.

在上面的简单数值例子中，我们只取小数点后两位数字. 在这一精度要求下，$Q = 2$，而且不论在哪一轮计算中，Q 都取相同值. 容易验证，当 $k \geqslant 3$ 时，$\hat{y}_{-k}^{(i)}$，$\hat{b}_{-k}^{(i)}$，$\hat{x}_{-k}^{(i)}$，$\hat{a}_{-k}^{(i)}$ 以及 $\hat{y}_{5+k}^{(i)}$，$\hat{x}_{5+k}^{(i)}$ 诸量的绝对值都比 0.005 小. 在这一例子中，只迭代两轮就得到了稳定值 $\hat{a}_k^{(2)}$，即 $|\hat{a}_k^{(2)} - \hat{a}_k^{(1)}| < 0.005$. 从而 ARMA(1, 1) 序列的平方和函数 $S(\boldsymbol{\beta})$，在样本值为 $\boldsymbol{w}_N^\tau = (-0.5, 0, 0.5, 0)$，参数 $\boldsymbol{\beta}^\tau = (0.3, 0.6)$ 时，有近似值

$$S(\boldsymbol{\beta}) \cong \sum_{k=-2}^{4} (\hat{a}_k^{(2)})^2 = (0.01)^2 + (0.03)^2 + (0.11)^2$$
$$+ (-0.46)^2 + (-0.13)^2 + (0.42)^2 + (0.10)^2 = 0.428$$

在实际应用中，若 N 较大时，不必从 $k = 0$ 开始计算 $\hat{b}_k^{(1)}$，可以从 $k = M$ (M 为远小于 N 的某个适当正整数) 开始计算，并且仍以 $\hat{b}_k^{(1)} = 0\ (k < M)$ 为初值. 一般说来，这时只迭代到 $\hat{a}_k^{(2)}(k \leqslant N)$ 便能得到稳定值.

§ 3. 精估计的数值解法及其它近似方法

前面所介绍的两种精估计方法，对于 MA 和 ARMA 模型来说都不能得到参数估计的明显表达式，只能用数值解法求 $S(\boldsymbol{\beta})$ (无论 (4.1.8) 或 (4.2.11) 式) 的极值解. 现在，将介绍几种求这类极值问题的数值方法.

表 4.2.1 ARMA(1, 1) 的平方和函数计算示例表

变量 / k	$\hat{y}_k^{(1)}$	$\hat{b}_k^{(1)}$	$\hat{x}_k^{(1)}$	$\hat{a}_k^{(1)}$	$\hat{y}_k^{(2)}$	$\hat{b}_k^{(2)}$	$\hat{x}_k^{(2)}$	$\hat{a}_k^{(2)}$	$\hat{y}_k^{(3)}$
⋮			⋮	⋮	⋮	⋮	⋮	⋮	
−3			(0)	(0) ⑥↓	(0)	(0) ①	(0)	(0) ⑤↓	
−2			0.01	0.01	−0.01	−0.01	0.01	0.01	
−1	⋮	⋮ 开始	0.03	0.03	−0.02	−0.02	0.03	0.03	
0	(0)	(0) ①↓	0.10 ↑④	0.11	−0.06 ↑⑧	−0.07	0.10 ↑④	0.11	
1	(0)	0.00	(−0.5)	−0.46	(0)	−0.02	(−0.5)	−0.46	
2	(0.5)	0.50	(0)	−0.13	(0.5)	0.49	(0)	−0.13	
3	(0)	0.15	(0.5)	0.42	(0)	0.14	(0.5)	0.42	
4	(−0.5)	−0.41	(0)	0.10	(−0.5)	−0.42	(0)	0.10	
5	0.10 ②↓	0	−0.06 ⑥↓	0	0.10 ②↓	0		停止	
6	0.03 ③↓	0	−0.02 ⑦↓	0	0.03 ⑧↓	0			
7	0.01	⋮	−0.01	⋮	0.01	⋮			
8	(0)		(0)		(0)				
⋮	⋮		⋮		⋮				

一、图解法

所谓图解法，就是在 β 的平稳可逆域内的网点上计算出 $S(\beta)$ 的值来，再通过绘图方法寻找 $S(\beta)$ 的极小值的近似解．这种方法与穷举法相似，计算量较大，但是容易掌握．在参数不多时（比如 $p+q=1$ 或 2）使用方便．我们只对 $p+q\leqslant 2$ 的情况分别具体叙述图解法．

$p+q=1$ 时的图解法计算步骤：

A．先求出矩估计 $\hat{\beta}^{(0)}$．

B．在 $\hat{\beta}^{(0)}$ 附近确定网格点 $\hat{\beta}^{(0)}\pm kh$（h 为步长，一般取 0.1 左右，当 N 很大时可适当取小些），计算 $S(\hat{\beta}^{(0)}\pm kh)$（$k=0,1,2,\cdots,k_0$）．

C．在坐标纸上将 $S(\hat{\beta}^{(0)}\pm kh)$ 绘出，并用光滑曲线拟合这些点值，然后寻找最低点值 $\hat{\beta}^{(1)}$．

D．再在 $\hat{\beta}^{(1)}$ 附近确定更密的格点 $\hat{\beta}^{(1)}\pm kh'$（$h'\ll h$）并计算 $S(\hat{\beta}^{(1)}\pm kh')$（$k=0,1,2,\cdots,k_1$）．

E．与C步相同，找出更精确的最小值 $S(\hat{\beta})$，则 $\hat{\beta}$ 为所求的近似解（可见下面例 1）．

$p+q=2$ 时的图解法计算步骤：

$p+q=2$ 的情况与上面类似，只是在绘图时不同．这时只能在参数平面的网格点上标出 $S(\beta)$ 的取值，然后勾画出等值线（见例 2），再找出最低值——“坑点”，与其相应的 $\hat{\beta}$ 为近似解．

当 $p+q>2$ 时，也可以采用图解法，不过计算量很大．比如 $p+q=3$ 时，可以固定第三个参数，其它两个变化，用 $p+q=2$ 的方法寻找最小值之解．然后将这两个变量固定在此解之处，让第三个变量变化，按 $p+q=1$ 时方法寻找最小值之解．反过来再求原来两个变量的新的极小化解．以此循环下去，直到三个解渐近稳定在某个值附近时便停止循环，并以最后的极小化解 $\hat{\beta}$ 作为图解法的近似解．由此可见，其计算量是很大的．最后，为了近一步熟悉图解方法，我们以（2.1.16 a），（2.1.16 e）式模型的最小二

乘估计为例，给出它们的图解法的解和某些分析说明.

例 1. 模型(2.1.16 a)式的参数最小二乘估计. 数据长度 $N = 300$，数据值与第三章 § 1，§ 3 所用的相同. 由于数据量太大没有一一列出. 现在用图解法求 θ_1 的最小二乘估计，此时由第四章 § 1 末尾的例子可知

$$S(\theta_1) = \sum_{k=1}^{300} \hat{a}_k^2 = \sum_{k=1}^{300}\left(\sum_{j=0}^{k-1} \theta_1^j w_{k-j}\right)^2$$

为了节省计算量，我们变换一下上式. 暂记 $w_k = 0$($k < 0$ 或 $k > 300$)，于是

$$\begin{aligned} S(\theta_1) &= \sum_{k=1}^{300}\left(\sum_{j=0}^{\infty} \theta_1^j w_{k-j}\right)^2 = \sum_{k=1}^{300}\left(\sum_{i,j=0}^{\infty} \theta_1^{i+j} w_{k-j} w_{k-i}\right) \\ &= 300 \sum_{i,j=0}^{\infty} \theta_1^i \theta_1^j \left(\frac{1}{300} \sum_{k=1}^{300} w_{k-j} w_{k-i}\right) \cong 300 \sum_{i,j=0}^{\infty} \theta_1^{i+j} \hat{\gamma}_{j-i} \\ &= 300\left(\sum_{i=0}^{\infty} \theta_1^{2i} \hat{\gamma}_0 + 2 \sum_{l=1}^{\infty} \theta_1^l \sum_{i=0}^{\infty} \theta_1^{2i} \hat{\gamma}_l\right) \\ &\cong 300\left(\frac{\hat{\gamma}_0}{1-\theta_1^2} + \frac{2}{1-\theta_1^2} \sum_{l=1}^{299} \theta_1^l \hat{\gamma}_l\right) \\ &\cong \frac{300\hat{\gamma}_0}{1-\theta_1^2}\left(1 + 2 \sum_{l=1}^{30} \theta_1^l \hat{\rho}_l\right) \end{aligned}$$

这样一来，就可以利用已有的计算结果 $\hat{\gamma}_0$ 和 $\hat{\rho}_k$($k = 1, 2, \cdots$)了. 再由矩估计可知 $\hat{\theta}_1^{(0)} = 0.38$，然后我们先将 $S(\theta_1)$ 在 $\theta_1 = 0.2$，0.3，0.4，0.5，0.6 的取值计算出来，描出 $S(\theta_1)$ 的曲线（见图 4.3.1），并发现最低点在 0.4 附近. 再将 $\theta_1 = 0.35$，0.45，0.55 处 $S(\theta_1)$ 的取值算出，描出新更细致的曲线，最后确定出 $\hat{\theta}_1 = 0.42$. 这就是我们所要的最小二乘估计的近似值. 它比第三章 § 3 的矩估计的精度高，此时误差仅为 -0.08，相应的 $\hat{\sigma}_a^2 = 1.06$，$\hat{\sigma}_a^2 = 0.06$，也比矩估计好.

例 2. 求(2.1.16 e)式的参数 (θ_1, φ_1) 的最小二乘估计. 与例 1 类似，$N = 300$；并且暂时记 $w_k = 0$($k < 0$ 或 $k > 300$). 依

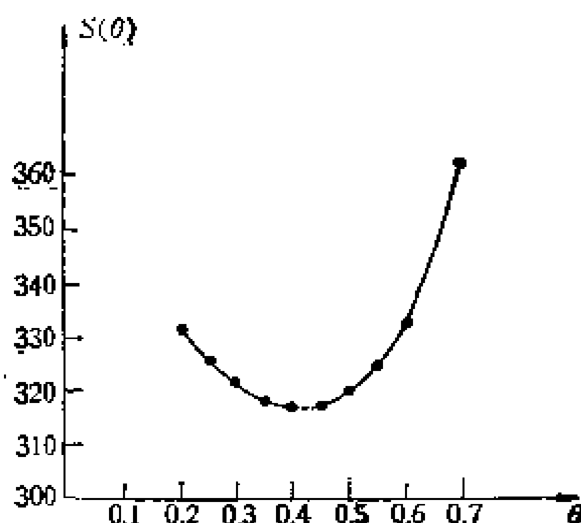


图 4.3.1

最小二乘估计方法的定义，

$$\hat{a}_k = w_k - \sum_{j=1}^{\infty} \pi_j w_{k-j} = (1 - \theta_1 B)^{-1}(1 - \varphi_1 B) w_k$$

$$= \sum_{j=0}^{\infty} \theta_1^j (w_{k-j} - \varphi_1 w_{k-j-1})$$

$$S(\boldsymbol{\beta}) = \sum_{k=2}^{300} \hat{a}_k^2 = \sum_{i,j=0}^{\infty} \theta_1^{i+j} \sum_{k=2}^{300}$$

$$\cdot (w_{k-j} - \varphi_1 w_{k-j-1})(w_{k-i} - \varphi_1 w_{k-i-1})$$

$$\cong \frac{300\hat{\gamma}_0}{1 - \theta_1^2}\Big\{1 + \varphi_1^2 - 2\varphi_1\hat{\rho}_1 + 2$$

$$\cdot \sum_{l=1}^{30} \theta_1^l [(1 + \varphi_1^2)\hat{\rho}_l - \varphi_1(\hat{\rho}_{l-1} + \hat{\rho}_{l+1})]\Big\}$$

再参考矩估计 $\hat{\theta}_1^{(0)} = 0.70$，$\hat{\varphi}_1^{(0)} = 0.56$，在 (θ_1, φ_1) 平面上先后定出粗网格点和细网格点．计算各网点上 $S(\boldsymbol{\beta})$ 的值，并绘出其等值线，找到坑点是 $\hat{\theta}_1 = 0.64$，$\hat{\varphi}_1 = 0.30$（见图 4.3.2）．它们与各自的真值的误差分别是 $\tilde{\theta}_1 = -0.04$，$\tilde{\varphi}_1 = 0.00$，可见比矩估计好．相应的 $\hat{\sigma}_a^2 = 1.06$，$\tilde{\sigma}_a^2 = 0.06$，这也比矩估计精确．

从以上两例还可看出，在参数 $\boldsymbol{\beta}$ 的维数不高（$\leqslant 2$），而 N 又较大的情况下，不必逐个求 $\hat{a}_k$，只通过 $\hat{\rho}_l$ 和 $\hat{\gamma}_0$ 便可近似计算出

$S(\boldsymbol{\beta})$来. 由于$|\theta|<1$，所以只用 30 个 $\hat{\rho}_l$ 值算出 $S(\boldsymbol{\beta})$ 的精度便很高了. 当然,在不同情况下 $\hat{\rho}_l$ 的个数有时多用些，有时少用些. 无论如何这种方法计算 $S(\boldsymbol{\beta})$ 的计算量远比 § 1 的办法小得多，而且方法简捷，这也是十分明显的. 关于这种近似方法的进一步推广，以及如何适当选取 $\hat{\rho}_l$ 的个数，使得既保证 $S(\boldsymbol{\beta})$ 的精度,又尽可能地减少计算量,这些问题还有待进一步研究.

至于用图解法求 (2.1.16 a)，(2.1.16 e) 的最小平方和估计，我

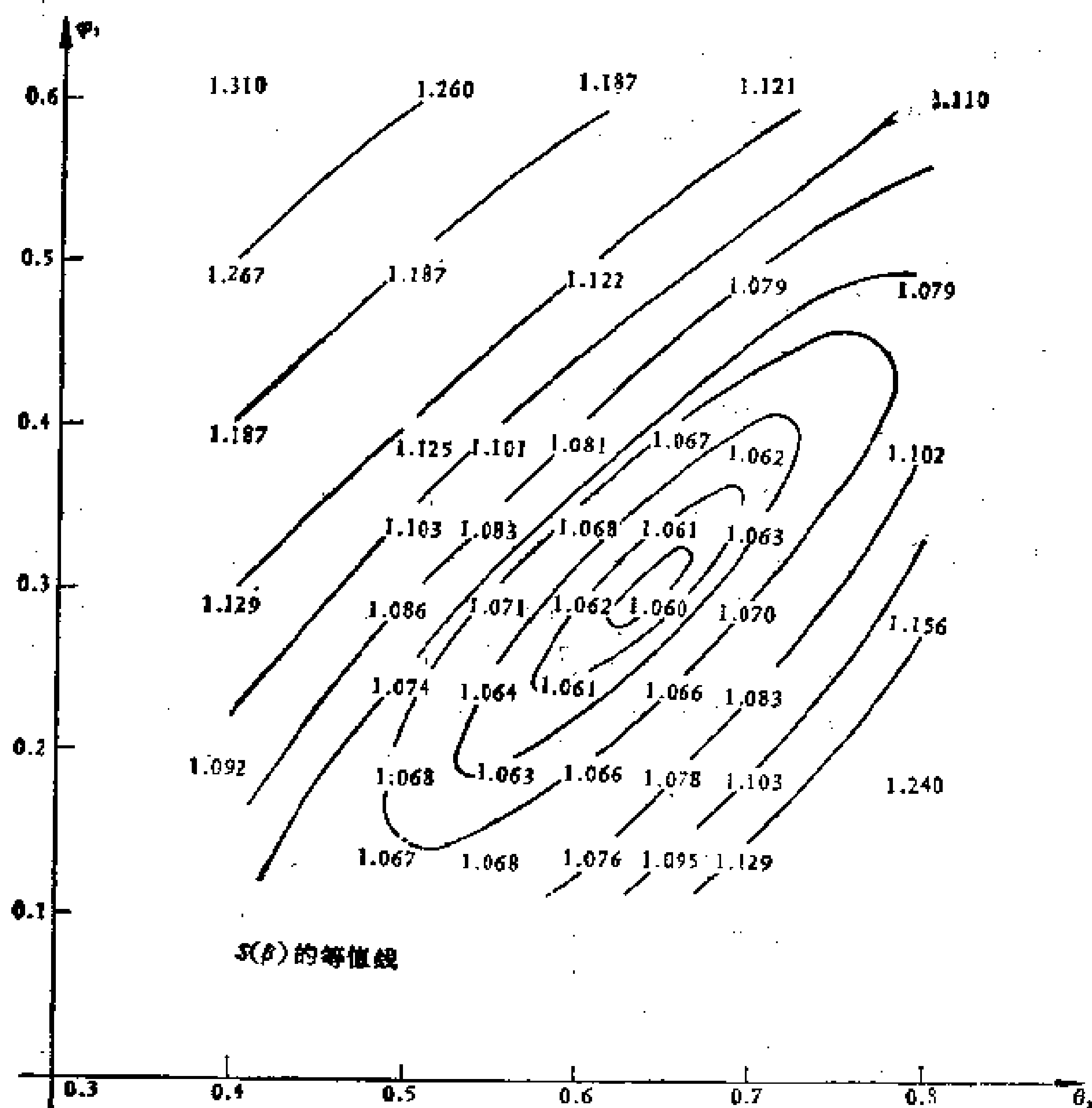


图 4.3.2

们不去列举了，那是十分烦琐的. 像上面 (2.1.16 a) 和 (2.1.16 c) 式这样的模型在其样本长度 $N=300$ 的情况下，最小平方和估计与最小二乘估计几乎没有差别，当然，使用例 1 、例 2 的方法是最方便的.

二、迭代法

1. 迭代法的一般叙述和步骤.

现在介绍用迭代法求 $S(\boldsymbol{\beta})$ 的极小值. 为了使得迭代收敛速度快一些，我们须要计算 $\hat{a}_t$ 的偏导数，即 $\hat{a}_t$ (或为 (4.1.7) 式或为 (4.2.1) 式所定义) 的梯度 $A_t=\frac{\partial \hat{a}_t}{\partial \boldsymbol{\beta}^\tau}$. 令 $r=p+q$, $\boldsymbol{\beta}^{(m)}$ 为 $\hat{\beta}$ 的第 m 步迭代值，以及

$$A_t^{(m)}=\frac{\partial \hat{a}_t}{\partial \boldsymbol{\beta}^\tau}\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}^{(m)}}=\left(\frac{\partial \hat{a}_t}{\partial \beta_1},\ \frac{\partial \hat{a}_t}{\partial \beta_2},\ \cdots,\ \frac{\partial \hat{a}_t}{\partial \beta_r}\right)\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}^{(m)}} \qquad (4.3.1)$$

根据台劳展开，$\hat{a}_t$ 在 $\boldsymbol{\beta}^{(m)}$ 附近可以展成

$$\hat{a}_t=\hat{a}_t^{(m)}+A_t^{(m)}(\boldsymbol{\beta}-\boldsymbol{\beta}^{(m)})+o(\boldsymbol{\beta}-\boldsymbol{\beta}^{(m)}),\quad -Q\leqslant t\leqslant N$$

即

$$\hat{a}_t^{(m)}=-A_t^{(m)}(\boldsymbol{\beta}-\boldsymbol{\beta}^{(m)})+\hat{a}_t+o(\boldsymbol{\beta}-\boldsymbol{\beta}^{(m)}),\quad -Q\leqslant t\leqslant N$$

其中 $o(\boldsymbol{\beta}-\boldsymbol{\beta}^{(m)})$ 表示比 $(\boldsymbol{\beta}-\boldsymbol{\beta}^{(m)})$ 为高阶小的向量. 将 $\hat{a}_t^{(m)}$ 视为第一章 § 5 中最小二乘模型的 y_t, $(\boldsymbol{\beta}-\boldsymbol{\beta}^{(m)})$ 视为待估计的参数，$A_t^{(m)}$ 为系数向量，$\hat{a}_t+o(\boldsymbol{\beta}-\boldsymbol{\beta}^{(m)})$ 为残量 e_t, 那么，按照最小二乘原理 $(\boldsymbol{\beta}-\boldsymbol{\beta}^{(m)})$ 可由下式估算 (根据 (1.5.3) 式)，即

$$\beta^*-\beta^{(m)}=-(H_m^\tau H_m)^{-1}H_m^\tau \boldsymbol{\xi}_m$$

其中

$$H_m=\begin{pmatrix} A_{-Q}^{(m)} \\ A_{-Q+1}^{(m)} \\ \vdots \\ A_N^{(m)} \end{pmatrix} \qquad \boldsymbol{\xi}_m=\begin{pmatrix} \hat{a}_{-Q}^{(m)} \\ \hat{a}_{-Q+1}^{(m)} \\ \vdots \\ \hat{a}_N^{(m)} \end{pmatrix} \qquad (4.3.2)$$

由最小二乘估计的定义可知，当 $\boldsymbol{\beta}=\boldsymbol{\beta}^*$ 时，最小二乘模型的残量平方和达到极小，而由于 $o(\boldsymbol{\beta}-\boldsymbol{\beta}^{(m)})$ 是高阶小量，所以上面所述的残量平方和能近似为

$$\sum_{t=-Q}^{N}\{a_t + o(\beta - \beta^{(m)})\}^2 \cong \sum_{t=-Q}^{N} a_t^2 = S(\beta)$$

我们的目的是选第 $m+1$ 步的迭代值 $\beta^{(m+1)}$，以使 $S(\beta^{(m+1)})$ 的值尽量小．根据上述分析，很自然地取

$$\beta^{(m+1)} = \beta^* = \beta^{(m)} - (H_m^\tau H_m)^{-1}H_m^\tau \xi_m \tag{4.3.3}$$

一般说来 $\beta^{(m+1)}$ 比 $\beta^{(m)}$ 更靠近真值 $\hat{\beta}$，这是由于

$$\xi_{m+1} \cong \xi_m + H_m(\beta^{(m+1)} - \beta^{(m)}) = \{I - H_m(H_m^\tau H_m)^{-1}H_m^\tau\}\xi_m$$

而矩阵 $\{I - H_m(H_m^\tau H_m)^{-1}H_m^\tau\}^\tau\{I - H_m(H_m^\tau H_m)^{-1}H_m^\tau\} = I - H_m(H_m^\tau H_m)^{-1}H_m^\tau \leqslant I$ 所以 $\xi_{m+1}^\tau\xi_{m+1} \leqslant \xi_m^\tau\xi_m$ 在近似的意义下成立，也就是说，在 $o(\beta - \beta^{(m)})$ 的确可忽略时，才有

$$\sum_{t=-Q}^{N}[a_t^{(m+1)}]^2 \leqslant \sum_{t=-Q}^{N}[a_t^{(m)}]^2$$

或者说，$S(\beta^{(m+1)})$ 一般比 $S(\beta^{(m)})$ 小一些．

综上所述迭代法的步骤如下：

A．根据 $\beta^{(m)}$ 计算 $a_t^{(m)}$ 和 H_m，当 $m=1$ 时，可取 β 的矩估计做为初值．关于 $a_t^{(m)}$ 的计算方法在本章 §2 第二小节中已经介绍了；至于 H_m 的计算方法下面将仔细介绍．

B．按照（4.3.3）式计算 $\beta^{(m+1)}$．

C．检查 $\|\beta^{(m+1)} - \beta^{(m)}\|$ 是否小于预定的精度界限．若小于时，则以 $\beta^{(m+1)}$ 做为 $\hat{\beta}$ 的近似值；否则再重复上述步骤，求出 $\beta^{(m+2)}$ 等．

表面上看来迭代法步骤简单，其实不然．步骤 A 所包括的计算，一类是求 $a_t^{(m)}$，一类是求 $\dfrac{\partial a_t^{(m)}}{\partial \beta^\tau}$，无论哪一类都要求出 $N+Q$ 个值来，计算量很大，计算公式十分复杂．计算 $a_t^{(m)}$ 的具体方法已经见到了，下面具体阐述求 $\dfrac{\partial a_t^{(m)}}{\partial \beta^\tau}$，即求 H_m 的方法．

2. H_m 的计算方法．

a．直接方法．用差分值近似微分值的方法求 $\left.\dfrac{\partial a_t}{\partial \beta_i^\tau}\right|_{\beta_i=\beta_i^{(m)}}$，令

$$\boldsymbol{\beta}^{(m)}(j) = (\beta_1^{(m)}, \beta_2^{(m)}, \cdots, \beta_{j-1}^{(m)}, \beta_j^{(m)} + \delta_j, \beta_{j+1}^{(m)}, \cdots, \beta_r^{(m)})$$

按照 $\hat{a}_t$ 的定义，计算 $\hat{a}_t^{(m)}(j) = \hat{a}_t(\boldsymbol{\beta}^{(m)}(j), \boldsymbol{w}_N)$，即以 $\boldsymbol{\beta}^{(m)}(j)$ 为参数值算出相应的 $\hat{a}_t$ 来 $(-Q \leqslant t \leqslant N)$. 然后取

$$\frac{\hat{a}_t^{(m)} - \hat{a}_t^{(m)}(j)}{\delta_j}$$

作为 $\left.\dfrac{\partial \hat{a}}{\partial \beta_j}\right|_{\beta_j=\beta_j^{(m)}}$ 的近似值. 对于每个 $j(=1, 2, \cdots, r)$ 都如此处理，便得到 $(N+Q)r$ 个偏导数的近似值，它们就构成了 H_m 的直接法估计值.

这一方法的差分步长 δ_j，依 N 的大小和精度要求而定. 直接法求 H_m 的计算量与求 r 次 $\hat{a}_t^{(m)}(-Q \leqslant t \leqslant N)$ 相差不多.

b. 迭代方法一. 这一方法只适用于求 $\boldsymbol{\beta}$ 的最小二乘估计中的 H_m. 根据本章 § 1 中的定义，

$$\hat{a}_k = w_k - \varphi_1 w_{k-1} - \cdots - \varphi_p w_{k-p} + \theta_1 \hat{a}_{k-1} + \cdots + \theta_q \hat{a}_{k-q}, \qquad p+1 \leqslant k \leqslant N \tag{4.3.4}$$

约定 $w_k = 0(k \leqslant 0)$, $\hat{a}_k = 0(k \leqslant 0)$. 上式可以写成

$$\hat{a}_k = \theta^{-1}(B)\varphi(B)w_k \tag{4.3.5}$$

利用此式分别求 $\hat{a}_k$ 的两种偏导数

$$\frac{\partial \hat{a}_k}{\partial \varphi_i} = -\theta^{-1}(B)B^i w_k = -\theta^{-1}(B)w_{k-i} = -\varphi^{-1}(B)\hat{a}_{k-i} \tag{4.3.6}$$

$$\frac{\partial \hat{a}_k}{\partial \theta_j} = \theta^{-2}(B)\varphi(B)B^j w_k = \theta^{-2}(B)\varphi(B)w_{k-j} = \theta^{-1}(B)\hat{a}_{k-j} \tag{4.3.7}$$

令 $h_t = -\varphi^{-1}(B)\hat{a}_t$, $g_t = \theta^{-1}(B)\hat{a}_t$，则

$$\frac{\partial \hat{a}_t}{\partial \varphi_i} = h_{t-i}, \qquad \frac{\partial \hat{a}_t}{\partial \theta_j} = g_{t-j} \tag{4.3.8}$$

再对 (4.3.4) 式两边求上述两种偏导数，并令 $\boldsymbol{\beta} = \boldsymbol{\beta}^{(m)}$，可得

$$w_{k-i} = -h_{k-i}^{(m)} + \theta_1^{(m)} h_{k-i-1}^{(m)} + \cdots + \theta_q^{(m)} h_{k-i-q}^{(m)} \tag{4.3.9}$$

$$o = g_{k-j}^{(m)} - \theta_1^{(m)} g_{k-j-1}^{(m)} - \cdots - \theta_q^{(m)} g_{k-j-q}^{(m)} - \hat{a}_{k-j}^{(m)} \tag{4.3.10}$$

其中 $h_{k-l}^{(m)}$ 和 $g_{k-l}^{(m)}$ 表示 h_{k-l} 和 g_{k-l} 在 $\boldsymbol{\beta}^{(m)}$ 处的值. 上式还可写成递推形式

$$h_k^{(m)} = \theta_1^{(m)} h_{k-1}^{(m)} + \theta_2^{(m)} h_{k-2}^{(m)} + \cdots + \theta_q^{(m)} h_{k-q}^{(m)} - w_k \tag{4.3.11}$$

$$g_k^{(m)} = \theta_1^{(m)} g_{k-1}^{(m)} + \theta_2 g_{k-2}^{(m)} + \cdots + \theta_q^{(m)} g_{k-q}^{(m)} + \hat{a}_k^{(m)} \tag{4.3.12}$$

仍然以零为初值，即取 $h_k^{(m)} = g_k^{(m)} = 0 (k \leqslant 0)$，于是 $h_k^{(m)}$ 和 $g_k^{(m)}$ $(0 < k \leqslant N)$ 便可由以上二式递推求得，其中 $\hat{a}_k^{(m)}$ 是通过本章 §2 第二小节的方法已经算好的第 m 步 $\hat{a}_t$ 的迭代值. 从而也就得到了 H_m 的值，

$$H_m^\tau = \{(A_1^{(m)})^\tau, (A_2^{(m)})^\tau, \cdots, (A_N^{(m)})^\tau\} \tag{4.3.13}$$

$$(A_k^{(m)}) = (h_{k-1}^{(m)}, h_{k-2}^{(m)}, \cdots, h_{k-p}^{(m)}, g_{k-1}^{(m)}, g_{k-2}^{(m)}, \cdots, g_{k-q}^{(m)})$$

用这个方法计算 H_m 时，其计算量不超过 $\hat{a}_k^{(m)} (p + 1 \leqslant k \leqslant N)$ 计算量的两倍；它不因 $r = p + q$ 的增加而成倍地增加.

c. 迭代方法二. 这一方法只适用于计算 $\boldsymbol{\beta}$ 的最小平方和估计的 H_m. 为求 H_m，只要求出每个 $\left.\dfrac{\partial \hat{a}_k}{\partial \boldsymbol{\beta}^\tau}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}^{(m)}}$ $(-Q \leqslant k \leqslant N)$. 完全可以按照本章 §2 第二小节求 $\hat{a}_k (-Q \leqslant k \leqslant N)$ 的方法建立 $\left.\dfrac{\partial \hat{a}_k}{\partial \boldsymbol{\beta}^\tau}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}^{(m)}}$ 的循环递推迭代公式，其基本思路是相同的，我们不详细论述. 仅以 ARMA(1, 1) 模型为例，介绍这种方法的具体迭代公式. 我们仍采用本章 §2 第二小节的记号. 有时为了简化符号，我们将 ARMA(1, 1) 的参数 φ_1, θ_1 的角标略去；对 AR(1, 0) 或 MA(0, 1) 亦有此情况. 此时由于 $\hat{x}_k = w_k$, $\hat{y}_k = w_{N-k+1}$ $(0 < k \leqslant N)$，所以 $\dfrac{\partial \hat{y}_k}{\partial \boldsymbol{\beta}} = 0$, $\dfrac{\partial \hat{x}_k}{\partial \boldsymbol{\beta}} = 0$ $(0 < k \leqslant N)$. 又由 (4.2.14) 和 (4.2.15) 式可知

$$\hat{a}_k = \hat{x}_k - \varphi \hat{x}_{k-1} + \theta \hat{a}_{k-1}, \qquad \hat{b}_k = \hat{y}_k - \varphi \hat{y}_{k-1} + \theta \hat{b}_{k-1}$$

若令 $\hat{a}_t^\varphi = \dfrac{\partial \hat{a}_t}{\partial \varphi}$, $\hat{a}_t^\theta = \dfrac{\partial \hat{a}_t}{\partial \theta}$，类似地还有 $\hat{x}_t^\varphi$, $\hat{x}_t^\theta$, $\hat{y}_t^\varphi$, $\hat{y}_t^\theta$，则

$$\hat{a}_k^\varphi = \hat{x}_k^\varphi - \varphi \hat{x}_{k-1}^\varphi + \theta \hat{a}_{k-1}^\varphi - \hat{x}_{k-1} \tag{4.3.14}$$

$$\hat{b}_k^\varphi = \hat{y}_k^\varphi - \varphi \hat{y}_{k-1}^\varphi + \theta \hat{b}_{k-1}^\varphi - \hat{y}_{k-1} \tag{4.3.15}$$

$$\hat{a}_k^\theta = \hat{x}_k^\theta - \varphi \hat{x}_{k-1}^\theta + \theta \hat{a}_{k-1}^\theta + \hat{a}_{k-1} \tag{4.3.16}$$

$$\hat{b}_k^\theta = \hat{y}_k^\theta - \varphi \hat{y}_{k-1}^\theta + \theta \hat{b}_{k-1}^\theta + \hat{b}_{k-1} \tag{4.3.17}$$

又由 (4.2.18) 和 (4.2.17) 式可知 $\hat{y}_k^\theta = \hat{x}_{N-k+1}^\theta$, $\hat{y}_k^\varphi = \hat{x}_{N-k+1}^\varphi$, 并且当 $k > N$ 时, $\hat{a}_k^\theta = \hat{b}_k^\theta = \hat{a}_k^\varphi = \hat{b}_k^\varphi = 0$.

若给出了 $\boldsymbol{\beta}$ (即 θ 和 φ) 之值，又求得了 $\hat{x}_k$ 和 $\hat{y}_k(-Q \leqslant k \leqslant N + Q + 1)$，并且 $\hat{x}_k \cong 0$，$\hat{y}_k \cong 0(k > N + Q + 1$ 或 $k < -Q)$，那么与建立 (4.2.26) 式到 (4.2.30) 式类似，根据以上诸式便可导出 $\hat{a}_k^\theta$ 和 $\hat{a}_k^\varphi$ 的递推迭代计算公式. 为简化记号暂时令 $\hat{x}_k^\varphi = v_k$，$\hat{y}_k^\varphi \equiv u_k$，$\hat{a}_k^\varphi \equiv l_k$，$\hat{b}_k^\varphi \equiv h_k$，于是易知这些量的初值分别为 $u_{-Q}^{(1)} = 0$ 和

$$u_k^{(i)} = v_k^{(i)} = l_k^{(i)} = h_k^{(i)} = 0, \quad k < -Q \text{ 或 } k > N + Q + 1$$

而它们的递推迭代公式可根据 (4.3.14) 和 (4.3.15) 两式得到，即

$$h_k^{(i)} = u_k^{(i)} - \varphi u_{k-1}^{(i)} + \theta h_{k-1}^{(i)} - \hat{y}_{k-1}, \quad -Q \leqslant k \leqslant N \tag{4.3.18}$$

$$\begin{cases} v_0^{(i)} = u_{N+1}^{(i)} = -\theta h_N^{(i)} + w_1 \\ v_{-1}^{(i)} = u_{N+2}^{(i)} = \varphi u_{N+1}^{(i)} + \hat{y}_{N+1} \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ v_{-Q}^{(i)} = u_{N+Q+1}^{(i)} = \varphi u_{N+Q}^{(i)} + \hat{y}_{N+Q} \end{cases} \tag{4.3.19}$$

$$l_k^{(i)} = v_k^{(i)} - \varphi v_{k-1}^{(i)} + \theta l_{k-1}^{(i)} - \hat{x}_{k-1}, \quad -Q \leqslant k \leqslant N \tag{4.3.20}$$

$$\begin{cases} u_0^{(i+1)} = v_{N+1}^{(i)} = -\theta l_N^{(i)} + w_N \\ u_{-1}^{(i+1)} = v_{N+2}^{(i)} = \varphi v_{N+1}^{(i)} + \hat{x}_{N+1} \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ u_{-Q}^{(i+1)} = v_{N+Q+1}^{(i)} = \varphi v_{N+Q}^{(i)} + \hat{x}_{N+Q} \end{cases} \tag{4.3.21}$$

$$h_k^{(i+1)} = u_k^{(i+1)} - \varphi u_{k-1}^{(i+1)} + \theta h_{k-1}^{(i+1)} - \hat{y}_{k-1}, \quad -Q \leqslant k \leqslant N$$

最后一式实际上又回到了 (4.3.18) 式. 这些公式与递推计算 $\hat{a}_k$ 的公式相比，差别在于这里多了自由项 $\hat{x}_k$ 和 $\hat{y}_k$，而这些量是在计算 $\hat{a}_k$ 时附带得到的数值结果. 对于 $\hat{a}_k^\theta$，也有类似上述的递推公式，我们不去仔细写出了.

总之，在用迭代法求解 $S(\boldsymbol{\beta})$ 的极小值时，当求出第 m 步的近似值 $\boldsymbol{\beta}^{(m)}$ 后，先要按 §2 第二小节的方法求出 $\hat{a}_k^{(m)}$ 和附带结果 $\hat{b}_k^{(m)}$, $\hat{x}_k^{(m)}$, $\hat{y}_k^{(m)}(-Q \leqslant k \leqslant N + Q + 1)$. 然后，用这些量和参数

$\beta^{(m)}$ 代替 (4.3.18) 至 (4.3.21) 式中的 $\tilde{a}_k$, $\tilde{y}_k$ 和 φ, θ, 经过反复循环迭代,便可求出 $\left.\frac{\partial \hat{a}}{\partial \varphi}\right|_{\varphi=\varphi^{(m)}}$ 的近似值. 求 $\left.\frac{\partial \hat{a}_k}{\partial \theta}\right|_{\theta=\theta^{(m)}}$ 也完全类似,只是在建立它们的迭代公式时用到的自由项是 $\hat{a}_k$ 和 $\hat{b}_k$, 而不是 $\tilde{a}_k$ 和 $\tilde{y}_k$; 因为这时使用的是 (4.3.16) 和 (4.3.17) 式. 从以上所述可知,求 H_m 的计算量与求 $\hat{a}_k(-Q \leqslant k \leqslant N)$ 的计算量的 $p+q$ 倍相差不多. 这种方法的精度较好. 但是,对于中等长度以上的样本数据(比如数百以上), 采用最小二乘法与最小平方和法没有多少差别,所以,只有在样本长度太小时才考虑使用最小平方和方法.

三、分离参数法

对于 ARMA 序列来说, 我们还要介绍两种介于矩估计与精估计之间的估计方法,它们既能减少计算的复杂程度,又能保证较好的精度.

1. 用矩估计方法分离参数 $\boldsymbol{\varphi}$. 当数据 $\boldsymbol{w}_N$ 的长度 N 较大时, 模型又被初步识别为 ARMA (p, q) 型, 我们按照第三章 § 3 的 (3.3.11) 式给出 $\boldsymbol{\varphi}$ 的矩估计 $\hat{\boldsymbol{\varphi}}$, 并由 (3.3.12) 式给出 $\bar{w}_k(p<k\leqslant N)$. 然后用这 $N-p$ 个 $\bar{w}_{p+1}, \bar{w}_{p+2}, \cdots, \bar{w}_N$ 为样本数据, 以 MA $(0, q)$ 为模型, 由精估计方法求相应的 $\boldsymbol{\theta}$ 的估计 $\hat{\boldsymbol{\theta}}$. 将 $\hat{\boldsymbol{\varphi}}$ 与 $\hat{\boldsymbol{\theta}}$ 合成 $\hat{\boldsymbol{\beta}}$, 就是一种分离的参数估计.

2. 建立只依赖 $\boldsymbol{\theta}$ 的平方和函数. 对于 ARMA 模型而言, $a_t = \varphi(B)\theta^{-1}(B)w_t$. 若令

$$\varepsilon_t = \theta^{-1}(B)w_t$$

则 ε_t 为一 AR 序列,而且满足关系式

$$\varepsilon_k = \theta_1\varepsilon_{k-1} + \theta_2\varepsilon_{k-2} + \cdots + \theta_q\varepsilon_{k-q} + w_k$$

若已给数据 $w_1, w_2, \cdots, w_N$ 和初值 $\varepsilon_{-q+1} = \varepsilon_{-q+2} = \cdots = \varepsilon_0 = 0$, 由上式可以推导出相应的 ε_t 的样本值 $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_N$. 若将它们视为 AR$(p, 0)$ 序列 $a_t = \varphi(B)\varepsilon_t$ 的样本值, 根据 § 2 第一小节,它的平方和函数的极小值为 (根据 (4.2.9) 式)

$$S(\hat{\varphi}_\theta) = d_{00} - \boldsymbol{d}^\tau D^{-1}\boldsymbol{d} \cong N(\hat{\gamma}_0(\theta) - \boldsymbol{b}^\tau(\theta)\hat{\Gamma}_p^{-1}(\theta)\boldsymbol{b}(\theta))$$

由于 $\varepsilon_k(1 \leqslant k \leqslant N)$ 与 $\boldsymbol{\theta}$ 有关，我们把 ε_t 的模型 $a_t = \varphi(B)\varepsilon_t$ 中的参数记为 $\boldsymbol{\varphi}_\theta$. 其中上式 $\hat{\boldsymbol{\varphi}}_\theta$ 表示对 $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_N$ 的模型参数$\boldsymbol{\varphi}_\theta$ 的最小平方和估计. 因此 $\boldsymbol{\varphi}_\theta$, $\hat{\boldsymbol{\varphi}}_\theta$ 以及 $\hat{\gamma}_0(\boldsymbol{\theta})$ 和 $\boldsymbol{b}^\tau(\boldsymbol{\theta}) = (\hat{\gamma}_1(\boldsymbol{\theta}), \hat{\gamma}_2(\theta), \cdots, \hat{\gamma}_p(\theta))$, $\hat{\Gamma}_p(\theta)$ 等都依赖于 $\boldsymbol{\theta}$, 同样 d_{00}, $\boldsymbol{d}$, D 也与 $\boldsymbol{\theta}$ 有关. $S(\hat{\boldsymbol{\varphi}}_\theta)$ 是依赖于 $\boldsymbol{\theta}$ 的 $\boldsymbol{w}_N$ 的平方和函数. 使 $S(\hat{\boldsymbol{\varphi}}_\theta)$ 达极小值的解 $\hat{\boldsymbol{\theta}}$, ,以及相应的 $\hat{\boldsymbol{\varphi}}_{\hat{\theta}}$ 和 $\frac{1}{N}S(\hat{\boldsymbol{\varphi}}_{\hat{\theta}})$ 就是 $\boldsymbol{\beta}$ 和 σ_a^2 的一种分离方法的参数估计.

例如考虑 ARMA$(1, q)$ 模型 $(1-\varphi(B))w_t = \theta(B)a_t$, 则

$$\varepsilon_k = \theta_1\varepsilon_{k-1} + \theta_2\varepsilon_{k-2} + \cdots + \theta_q\varepsilon_{k-q} + w_k$$

以 $\varepsilon_{-q+1} = \varepsilon_{-q+2} = \cdots = \varepsilon_0 = 0$ 为初值，对每个给定的$\boldsymbol{\theta}$ 值便得到 N 个 ε_k(依赖于 $\boldsymbol{\theta}$) 之值. 用这N个数据求出 φ 的最小平方和估计为

$$\hat{\varphi}_\theta = \hat{\rho}(\theta) = \left(\sum_{k=2}^{N} \varepsilon_k\varepsilon_{k-1}\right)\left(\sum_{k=2}^{N-1} \varepsilon_k^2\right)^{-1}$$

$$\begin{aligned}\hat{\sigma}_a^2(\boldsymbol{\theta}) &= \frac{1}{N}\sum_{k=1}^{N} \varepsilon_k^2 - \frac{1}{N}\left(\sum_{k=2}^{N} \varepsilon_k\varepsilon_{k-1}\right)^2\left(\sum_{k=2}^{N-1} \varepsilon_k^2\right)^{-1} \\ &\cong \hat{\gamma}_0(\boldsymbol{\theta}) - \frac{\hat{\gamma}_1^2(\boldsymbol{\theta})}{\hat{\gamma}_0(\boldsymbol{\theta})}\end{aligned}$$

在所有可能的 $\boldsymbol{\theta}$ 值中 (即在可逆域内), 选出使上式 $\hat{\sigma}_a^2(\boldsymbol{\theta})$ 达到极小值的 $\hat{\boldsymbol{\theta}}$, 便可得到 $\boldsymbol{\beta}$ 和 σ_a^2 的估计. 这种方法减少了最小平方和估计计算的复杂程度,而且精度也能保持相同,不过求 $\hat{\sigma}_a^2(\boldsymbol{\theta})$ 的极值问题仍须用数值解法.

§4. 极大似然估计与估计的优效性质

一、似然函数与极大似然估计

仍以 $\boldsymbol{w}_N$ 表示正态 ARMA (p, q) 序列的一段取样，它就是N维正态随机向量的样本. 用 $p(\boldsymbol{w}_N|\boldsymbol{\beta}, \sigma_a^2)$ 表示它的分布密度，其中 $(\boldsymbol{\beta}^\tau, \sigma_a^2) = \boldsymbol{\alpha}^\tau$ 表示 w_t 的全部参数. 依第一章(1.2.3)式可知

$p(\boldsymbol{w}_N|\boldsymbol{\beta}, \sigma_a^2)$ 被 w_t 的自协方差函数唯一决定．又由第二章可知，w_t 的自协方差函数由 $\boldsymbol{\beta}, \sigma_a^2$ 唯一决定．可见，当 $\boldsymbol{\varphi}$ 和 $\boldsymbol{\theta}$ 在平稳可逆域中变化，σ_a^2 在 $(0, \infty)$ 内变化时，$p(\boldsymbol{w}_N|\boldsymbol{\beta}, \sigma_a^2)$ 就构成一个特殊的 N 维正态分布密度函数族，我们用 $\mathscr{F}_{p,q}$ 表示．给了一组参数值 $\boldsymbol{\alpha}$（即 $\boldsymbol{\varphi}, \boldsymbol{\theta}$ 和 σ_a^2），便可在 $\mathscr{F}_{p,q}$ 中找到一个相应的分布密度函数 $p(\boldsymbol{w}_N|\boldsymbol{\beta}, \sigma_a^2)$，或记做 $p(\boldsymbol{w}_N|\boldsymbol{\alpha})$．

当得到 $\boldsymbol{w}_N$ 的一个样本，即得到由 N 个数据组成的序列，我们由模型识别判断它为 ARMA(p, q) 序列的一段样本值，就等于判断出 $\boldsymbol{w}_N$ 的分布密度是属于族 $\mathscr{F}_{p,q}$．前面所介绍的参数估计，就是要进一步确定 $\boldsymbol{w}_N$ 的分布函数是族 $\mathscr{F}_{p,q}$ 中的那一个，也就是要把决定它的参数 $\boldsymbol{\alpha}$ 估计出来．根据分布和分布密度的定义，$p(\boldsymbol{w}_N|\boldsymbol{\alpha})$ 描述了当 w_t 的分布以 $\boldsymbol{\alpha}$ 为参数时，样本值 $\boldsymbol{w}_N$ 出现的可能性的大小．所以根据上面的叙述，当我们由取得的具体样本值 $\boldsymbol{w}_N$ 定出了它的模型为 ARMA(p, q)，我们再对 $\boldsymbol{\alpha}$ 极大化 $p(\boldsymbol{w}_N|\boldsymbol{\alpha})$，令其最大的一个记做 $p(\boldsymbol{w}_N|\hat{\boldsymbol{\alpha}})$．这说明当 w_t 的分布以 $\hat{\boldsymbol{\alpha}}$ 为参数时，出现这段数据 $\boldsymbol{w}_N$ 的可能性比 $\boldsymbol{\alpha}$ 为其它值时都要大．自然，可把这种方法所确定的 $\hat{\boldsymbol{\alpha}}$ 称为 $\boldsymbol{\alpha}$ 的极大概率密度估计．为了求 $p(\boldsymbol{w}_N|\boldsymbol{\alpha})$ 的极大值，我们把 $\boldsymbol{w}_N$ 当做固定的参数量，而把 $\boldsymbol{\alpha}$ 看做 $r+1(=p+q+1)$ 维的自变量，为了突出表示这种函数概念，特别令

$$L(\boldsymbol{\alpha}|\boldsymbol{w}_N) = p(\boldsymbol{w}_N|\boldsymbol{\alpha}) \tag{4.4.1}$$

称 L 为 $\boldsymbol{w}_N$ 的似然函数．上面所说的极大概率密度估计，就变成求 $L(\boldsymbol{\alpha}|\boldsymbol{w}_N)$ 达极大值的问题了．因此，一般都称上述的 $\hat{\boldsymbol{\alpha}}$ 为极大似然估计，而不用极大概率密度的名称．对于正态分布来说，$p(\boldsymbol{w}_N|\boldsymbol{\alpha})$ 是指数型函数，在求解极值问题时，采用以下等价的求极值问题更为方便，即令

$$l(\boldsymbol{\alpha}|\boldsymbol{w}_N) = \log L(\boldsymbol{\alpha}|\boldsymbol{w}_N) \tag{4.4.2}$$

l 称为 $\boldsymbol{w}_N$ 的对数似然函数．于是求极大似然估计的问题又化成求解以下方程的问题，即

$$\left.\frac{\partial l}{\partial \boldsymbol{\alpha}}\right|_{\alpha=\hat{\alpha}} = 0 \tag{4.4.3}$$

现在，我们举一个十分简单的例子．设 $w_t = a_t$ 是一白噪声，或者称为 ARMA(0，0) 序列．如果我们获得了 N 个数据 $\boldsymbol{w}_N$，那么

$$p(\boldsymbol{w}_N|\sigma_a^2) = (2\pi\sigma_a^2)^{-\frac{N}{2}} e^{-\frac{1}{2\sigma_a^2}(w_1^2+w_2^2+\cdots+w_N^2)}$$

$$l(\sigma_a^2|\boldsymbol{w}_N) = -\frac{N}{2}\log 2\pi - \frac{N}{2}\log \sigma_a^2 - \frac{1}{2\sigma_a^2}(w_1^2 + w_2^2 + \cdots + w_N^2)$$

由

$$\frac{\partial l}{\partial \sigma_a^2} = -\frac{N}{2\sigma_a^2} + \frac{w_1^2 + w_2^2 + \cdots + w_N^2}{2\sigma_a^4} = 0$$

解出 $\hat{\sigma}_a^2 = \frac{1}{N}(w_1^2 + w_2^2 + \cdots + w_N^2)$．这就是 σ_a^2 的极大似然估计．

对于一般的 ARMA(p, q) 序列，其参数的似然函数及极大似然估计要复杂得多．下面仅做简单的讨论，某些烦琐的论证放在附录中．

设 w_t 为正态 ARMA 序列，其 $\boldsymbol{w}_N$ 的分布密度函数为（见(1.2.3)式）

$$p(\boldsymbol{w}_N|\boldsymbol{\alpha}) = \frac{1}{(2\pi)^{\frac{N}{2}}|\Gamma_N|^{\frac{1}{2}}} e^{-\frac{\boldsymbol{w}_N^{\tau}\Gamma_N^{-1}\boldsymbol{w}_N}{2}} = \frac{|M_N|^{\frac{1}{2}}}{(2\pi\sigma_a^2)^{\frac{N}{2}}} e^{-\frac{\boldsymbol{w}_N^{\tau}M_N\boldsymbol{w}_N}{2\sigma_a^2}} \tag{4.4.4}$$

其中 M_N 的定义见(4.2.4)式．再用(4.2.3)式，$\boldsymbol{w}_N$ 的对数似然函数可以写成

$$l(\boldsymbol{\alpha}|\boldsymbol{w}_N) = -\frac{N}{2}\log 2\pi + \frac{1}{2}\log|M_N| - \frac{N}{2}\log\sigma_a^2 - \frac{S(\boldsymbol{\beta})}{2\sigma_a^2} \tag{4.4.5}$$

上式右边的第一项与 $\boldsymbol{\alpha}$ 无关．在附录§3中还证明了，$\log|M_N|$ 有与 N 无关的上界．而上式右边的后两项是与 N 同阶的

量．所以，当N较大时，为了求$l(\boldsymbol{\alpha}|\boldsymbol{w}_N)$的极大值，可以用求后两项的极大值来近似，即(4.4.3)式的解可由以下公式的解来近似，

$$\frac{\partial}{\partial\sigma_a^2}\left\{-\frac{N}{2}\log\sigma_a^2-\frac{S(\boldsymbol{\beta})}{2\sigma_a^2}\right\}\Bigg|_{\alpha=\hat{\alpha}}=\frac{-N}{2\hat{\sigma}_a^2}+\frac{S(\hat{\boldsymbol{\beta}})}{2\hat{\sigma}_a^4}=0 \quad (4.4.6)$$

$$\frac{\partial}{\partial\boldsymbol{\beta}^\tau}\left\{-\frac{N}{2}\log\sigma_a^2-\frac{S(\boldsymbol{\beta})}{2\sigma_a^2}\right\}\Bigg|_{\alpha=\hat{\alpha}}=-\frac{1}{2\hat{\sigma}_a^2}\frac{\partial S(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}^\tau}\Bigg|_{\beta=\hat{\beta}}=0 \quad (4.4.7)$$

也就是说，$\hat{\sigma}_a^2=\frac{1}{N}S(\hat{\boldsymbol{\beta}})$，而$\hat{\boldsymbol{\beta}}$是$\frac{\partial S(\boldsymbol{\beta})}{\partial\beta}=0$的解．这正是本章§2中的最小平方和估计．它是极大似然估计的很好的近似解，所以我们又称它为近似极大似然估计．

严格极大似然估计的求解一般也只能用数值解法．因为解法更加烦琐，通常很少采用它．在数据很少而又希望尽可能地提高估计精度情况下，偶而也用到这种估计．例如，某种随机现象作为过程被考查时，每次观测只能得到少量的数据．但是，这种现象却可以重复观测．也就是说，能够获得多个相互独立的样本数据$\boldsymbol{w}_N^{(k)}(k=1,2,\cdots,l)$．每次样本长度$N$不太大，而每个$\boldsymbol{w}_N^{(k)}$都是对同一随机过程的独立观测的数据．我们希望对每个$\boldsymbol{w}_N^{(k)}$进行尽可能细致的分析估计，然后再从这l组估计决定出总的估计结果．这时就要考虑使用严格极大似然估计方法了．如果仅有一组$\boldsymbol{w}_N$，同时N又很小，那么即使用严格极大似然估计，也不会得到很精确的结果．因为样本长度太小，其信息量就少．所谓更精细的估计方法，只是相对于粗的或较精的估计而言的；并不能指望用它来克服由于样本小所带来的估计误差大的困难．

为了计算严格极大似然估计，还希望把ARMA序列的对数似然函数写得更具体一些，读者可以查看本书附录§3，那里给出了各种序列的分布密度的表达式．

二、估计的优效性质

到这里为止，我们在本书中已经提到了很多种估计量了，比如$\hat{\gamma}_k$，$\hat{\rho}_k$，$\hat{\varphi}_{kk}$以及$\hat{\beta}$，$\hat{\sigma}_a^2$等．统计上所谓的统计量，不外是某些随

机变量或随机序列的样本的函数．给出这种函数的方法称之为估计方法，例如最小二乘法，最小平方和法，极大似然法，以及很多没有命名的方法．在第一章§5第三小节中，我们介绍了衡量估计量优劣程度的几种优效性概念．在这一节里，我们讨论ARMA序列的参数估计的优效性质．

研究ARMA序列参数的各种估计的优效性质，既有理论价值又有实际意义，不少文章（如[9][15][16][19]）提供了极大似然估计的某些优效性质．但是，如在第一章§5所指出的那样，由于某些统计著作中在渐近优效性这一概念上的混乱，即把本书中所称的优效渐近正态性与渐近优效性（实质上是渐近最小方差性）混同起来，也造成了时间序列分析中参数估计的渐近优效性概念的混乱和谬误．产生这种混乱现象的另一方面的原因，是由于估算 $\tilde{\alpha}=\hat{\alpha}-\alpha$ 的方差阵 $E\tilde{\alpha}\tilde{\alpha}'$ 时遇到了困难．有的文章（如[9]）为了迴避估算 $E\tilde{\alpha}\tilde{\alpha}'$，改换渐近优效性的定义，而代之以本书中所称的优效渐近正态性，但是却保留原来的名称．有的文章（如[1]）虽然仍用原来定义的渐近优效性，但其论证却有错误．有的文章（如[4]）还引入了新的估计概念，以绕过计算 $E\tilde{\alpha}\tilde{\alpha}'$ 所带来的困难．为了澄清这种混乱，我们明确渐近优效性和优效渐近正态性为两种不同的渐近性质，分别给出定义（见第一章§5第三小节）．就我们所了解的情况而言，文献中较多地讨论了ARMA序列参数估计的相容性，以及渐近正态性．至于第一章§5第三小节明确定义的渐近优效性，虽有文章涉及到，但是没有见到严格的讨论．

在本书附录（主要是附录§5）中，我们证明了，对于正态AR(p, 0)序列的参数估计（无论本书介绍的那一种），它们具有第一章§5第三小节所定义的各种渐近性质，其中包括首次给出严格证明的渐近优效性，即

(a) 渐近无偏性：$\lim\limits_{N\to\infty}E\hat{\alpha}=\alpha^0$;

(b) 相容性：$\lim\limits_{N\to\infty}\hat{\alpha}\overset{p}{=}\alpha^0$;

(c) 渐近优效性: $E\tilde{\boldsymbol{\alpha}}\tilde{\boldsymbol{\alpha}}^{\tau}=\left(E\frac{\partial l}{\partial \boldsymbol{\alpha}}\frac{\partial l}{\partial \boldsymbol{\alpha}^{\tau}}\right)^{-1}+o(N^{-1})$;

(d) 渐近正态性: $\sqrt{N}\tilde{\boldsymbol{\alpha}}$ 的分布渐近正态;

(e) 优效渐近正态性: $\left(E\frac{\partial l}{\partial \boldsymbol{\alpha}}\frac{\partial l}{\partial \boldsymbol{\alpha}^{\tau}}\right)^{\frac{1}{2}}\tilde{\boldsymbol{\alpha}}$ 的分布渐近正态 $\mathscr{N}(0, I)$.

这一结果推广到非正态的 AR 序列情况并不困难. 至于对一般的正态 ARMA 序列以及非正态的情形，是否仍有上述结果的问题，前面已经指出，除性质 (c) 此外，已有人进行了不同程度的研究. 此外，最近的文章，对非正态多维 AR 序列的参数估计的渐近优效性，进行了较细致的研究. 总之，在渐近性理论方面的工作还有待进一步的深入研究.

从 ARMA 序列的参数估计的渐近性质中可以看到，估计误差 $\tilde{\boldsymbol{\alpha}}$ 的方差阵 $E\tilde{\boldsymbol{\alpha}}\tilde{\boldsymbol{\alpha}}^{\tau}$ 的 Cramer-Rao 下界是很有用的. 在近代的时序分析文献中，常把它称做信息矩阵，并记做

$$J_N=E\frac{\partial l}{\partial \boldsymbol{\alpha}}\frac{\partial l}{\partial \boldsymbol{\alpha}^{\tau}}=-E\frac{\partial^2 l}{\partial \boldsymbol{\alpha}\,\partial \boldsymbol{\alpha}^{\tau}} \tag{4.4.8}$$

其中第二个等式是在一定的条件下（如正态序列）才成立的，我们在附录 § 4 中将给出较宽的条件，并在该条件下证明此式. 并且，还证明了 AR$(p, 0)$ 序列的信息矩阵有渐近式:

$$J_N(\boldsymbol{\alpha})\approx\frac{N}{\sigma_a^2}\begin{pmatrix}\Gamma_p & 0\\ 0 & \frac{1}{2\sigma_a^2}\end{pmatrix}=N\begin{pmatrix}M_p^{-1} & 0\\ 0 & \frac{1}{2\sigma_a^4}\end{pmatrix} \tag{4.4.9}$$

ARMA(p, q) 序列的信息矩阵有渐近式:

$$J_N(\boldsymbol{\alpha})\approx\begin{pmatrix}I_p & 0\\ 0 & -I_{q+1}\end{pmatrix}J_N^{\mathrm{AR}}(\boldsymbol{\alpha})\begin{pmatrix}I_p & 0\\ 0 & -I_{q+1}\end{pmatrix} \tag{4.4.10}$$

其中 $J_N^{\mathrm{AR}}(\boldsymbol{\alpha})$ 为如下 AR$(p+q, 0)$ 序列 w_t 的信息矩阵，即

$$\varphi(B)\theta(B)w_t=a_t \tag{4.4.11}$$

最后，列举几个常用的低阶 ARMA 序列信息矩阵之逆的渐近表达式:

AR(1, 0): $J_N^{-1}(\varphi, \sigma_a^2) \approx \frac{1}{N}\begin{pmatrix} 1-\varphi^2 & 0 \\ 0 & 2\sigma_a^4 \end{pmatrix}$

AR(2, 0):

$$J_N^{-1}(\varphi_1, \varphi_2, \sigma_a^2) \approx \frac{1}{N}\begin{pmatrix} 1-\varphi_2^2 & -\varphi_1(1+\varphi_2) & 0 \\ -\varphi_1(1+\varphi_2) & 1-\varphi_2^2 & 0 \\ 0 & 0 & 2\sigma_a^4 \end{pmatrix}$$

MA(0, 1): $J_N^{-1}(\theta, \sigma_a^2) \approx \frac{1}{N}\begin{pmatrix} 1-\theta^2 & 0 \\ 0 & 2\sigma_a^4 \end{pmatrix}$

MA(0, 2):

$$J_N^{-1}(\theta_1, \theta_2, \sigma_a^2) \approx \frac{1}{N}\begin{pmatrix} 1-\theta_2^2 & -\theta_1(1+\theta_2) & 0 \\ -\theta_1(1+\theta_2) & 1-\theta_2^2 & 0 \\ 0 & 0 & 2\sigma_a^4 \end{pmatrix}$$

ARMA(1, 1):

$$J_N^{-1}(\varphi, \theta, \sigma_a^2) \approx \frac{1}{N}$$

$$\cdot\begin{pmatrix} \dfrac{(1-\varphi^2)(1-\varphi\theta)^2}{(\varphi-\theta)^2} & \dfrac{(1-\varphi^2)(1-\theta^2)(1-\varphi\theta)}{(\varphi-\theta)^2} & 0 \\ \dfrac{(1-\varphi^2)(1-\theta^2)(1-\varphi\theta)}{(\varphi-\theta)^2} & \dfrac{(1-\theta^2)(1-\varphi\theta)^2}{(\varphi-\theta)^2} & 0 \\ 0 & 0 & 2\sigma_a^4 \end{pmatrix}$$

在实际应用中，根据这些公式可以了解和比较所得到的估计值的误差情况．例如，考虑第三章§3中模型（2.1.16 e）的矩估计的误差，此时模型（2.1.16 e）的信息矩阵之逆为

$$J_N^{-1}(\varphi, \theta, \sigma_a^2) \approx \frac{1}{300}\begin{pmatrix} 6.8 & 5.3 & 0 \\ 5.3 & 4.7 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

（注意这里用了真值 $\varphi^0 = 0.3, \theta^0 = 0.6, \sigma_a^2 = 1$．）由此可见，第三章§3表3.2.1中(2.1.16e)的 $\bar{\varphi} = 0.26$，与 $E\bar{\varphi}^2$ 的 Cramer-Rao 下界的平方根 $\sqrt{\frac{6.8}{300}} \cong 0.15$ 相比，超过了 0.15，但仍小于二倍的

0.15. 这时,不能认为这个估计是很好的. 而在本章§3的例2中 φ 的最小二乘估计为0.3, 相应的 $\hat{\varphi}=0.00$, 这是一个非常好的估计. 但是,在实际应用中信息矩阵所依赖的真实参数是未知的,因此,只能用它们的估值来代替,以此来近似了解估计的误差方差大小.

在使用上述公式时,还须注意两点. 其一,以上诸渐近公式只在 $\boldsymbol{\beta}$ 远离平稳可逆域边界时才能有效使用; 其二, $\varphi(B)$, $\theta(B)$ 的根相差比较大时才能使用 ARMA(p, q) 的信息矩阵的渐近公式. 若 $\varphi(B)$ 与 $\theta(B)$ 有相等的根, 则信息矩阵的逆矩阵会出现发散现象,即它的个别元变得无限大. 例如, 在 ARMA(1, 1) 序列中,若 $\varphi=\theta$, 则 w_t 实际上退化为白噪声. 这时模型参数只要满足 $\varphi=\theta$, 它们可以取任意值, 因此对它们的估计便成为无意义的事了. 这样的参数称为"多余参数". 所以我们在定义 ARMA 序列时,总假定 $\varphi(B)$ 与 $\theta(B)$ 无公共因子. 但是,如果它们的某个根十分相近时,仍会产生上述的发散现象; 在实际应用中,如果出现这种现象, 可以降低估计模型的阶数, 以消除 $\varphi(B)$ 与 $\theta(B)$ 的相近的根.

第五章　估计模型的检验与改进

§1. 估计模型检验的一般叙述

客观世界中出现的随机序列是纯粹的 ARMA 序列者并不多见．用 ARMA 序列来描述实际中的随机序列，主要是当做一种近似手段使用．根据已获得的数据，我们希望用这种描述和估计能反映出真实随机序列的基本统计特性，以便进一步在预报、控制或模拟等课题中运用这些特性．所以前两章讨论的识别和估计的过程，可看做是选用适当的 ARIMA(p, d, q) 模型（可称为估计模型）来拟合真实随机序列．至于拟合的优劣程度如何，主要通过实际应用效果来检验．在将一种估计模型用于实际之前，最好先用数学方法对估计模型做一番检验，看看它与真实情况的相近程度如何．

鉴别一种估计模型的好坏，其方法也是多种多样的，可以根据实际问题的不同要求采用不同的方法．比如，若建立估计模型仅仅是为了进行预报的目的，这时，常按估计模型对已有的数据进行预报(具体方法见第七章)，然后把预报误差的平方和再加以平均，一般说来，这个量越小，估计模型与真实序列就符合得越好．此外，还有所谓的"过拟合"法，就是增大模型阶数，根据样本数据重新估计各参数值，若新加入的高阶项系数的估计值很小，就认为原来的估计模型是合适的．这种方法的计算量较大，并不常用，只对 AR 模型比较方便．

本章将介绍一些统计检验方法，以鉴别估计模型能否被接受，这些方法用起来方便，适用范围比较广．其主要原理是统计学中的假设检验理论（见 [6]）．进行检验的基本想法叙述如下．

假定 z_t 被估计为 ARIMA(p, d, q) 序列，即 $\varphi(B)\nabla^d z_t = \theta(B)a_t$，那么 $a_t = \theta^{-1}(B)\varphi(B)\nabla^d z_t$ 就应当近似为白噪声序列．

因此，若能从 z_t 的一段样本求得 a_t 的一段样本值 $a_1, a_2, \cdots, a_N$，便可以对"a_t 是白噪声序列"这一命题进行统计中的假设检验．如果肯定这一命题，就认为模型拟合得较好；若否定这一命题，则模型拟合得不好．在假设检验中，显著性水平 α 常用0.01，0.05或0.1 等值．这里包含两个重要问题：首先，要检验 a_t 是否是白噪声有哪些可供采用的方法？关于这一点我们将在本章的 § 2，§ 3 两节中分别介绍两种检验方法，也还有其他方法可用（如[5]）．其次，怎样获得 a_t 的一段样本值？本节先讨论第二个问题．

假定对被考查的随机序列 z_t 得到的估计模型是

$$\varphi(B)\nabla^d z_t = \theta(B)a_t \tag{5.1.1}$$

由于本节重点是讨论模型检验问题，所以 $\varphi(B)$，$\theta(B)$ 的诸参数仍以 $\boldsymbol{\varphi}$ 和 $\boldsymbol{\theta}$ 表示（σ_a^2 也如此），而不再用 $\hat{\phi}$，$\hat{\theta}$了．此外，还假定原序列有一段样本 $z_1, z_2, \cdots, z_N$，它可以是原来用以识别估计模型时的数据，也可以是另外一串新的数据(但仍是取自同一随机序列的样本值)．我们要从这串数据和 (5.1.1) 式求得 a_t的一段样本值．记 $\nabla^d z_t = w_t (t > d)$，则 (5.1.1) 式可写成

$$\varphi(B)w_t = \theta(B)a_t \tag{5.1.2}$$

其中依约定，w_t 是均值为零并与 $z_1, \cdots, z_d$ 独立的 ARMA(p, q) 序列．样本值 $w_{d+1}, w_{d+2}, \cdots, w_N$ 可以从 $z_1, z_2, \cdots, z_N$ 获得，有时也把后者称为"量测数据"．从 (5.1.2) 式可以得到

$$a_t = \theta^{-1}(B)\varphi(B)w_t = \pi(B)w_t \tag{5.1.3}$$

a_t 称为用 (5.1.2) 式拟合序列 w_t (或用 (5.1.1) 式拟合 z_t) 的"残量"．它的样本值又可以通过 w_t 的样本值求得．由于 $\{\pi_j\}$ 下降速度与负指数律同阶，(5.1.3) 式的无穷和可以用有穷和近似．于是当已知 $\boldsymbol{w}_{N-d} = (w_{d+1}, w_{d+2}, \cdots, w_N)^\tau$ 时，

$$a_k \cong w_k - \sum_{j=1}^{k-1} \pi_j w_{k-j}, \quad k = Q+1, Q+2, \cdots, N \tag{5.1.4}$$

其中 Q 应取得适当大，以使 $\sum\limits_{j=Q}^{\infty} \pi_j$ 为可忽略的量．当 (5.1.2) 式

为纯 AR 型,即 $\varphi(B)w_t = a_t$ 时,则

$$a_k = w_k - \varphi_1 w_{k-1} - \varphi_2 w_{k-2} - \cdots - \varphi_p w_{k-p}$$
$$k = Q+1, Q+2, \cdots, N \tag{5.1.5}$$

其中取 $Q = p + d$. 这就是说，当估计模型为 AR 型时，从量测数据 $z_1, z_2, \cdots, z_N$ 可以得到准确的、相应的“残量”样本值 a_{Q+1}, $a_{Q+2}, \cdots, a_N$. 对一般的 ARMA 模型,为了提高 $a_k(N \geqslant k \geqslant Q)$ 的近似精度,采用以下的公式

$$a_k \cong w_k - \sum_{j=1}^{k-1} \pi_j w_{k-j} - \sum_{j=k}^{\infty} \pi_j \hat{w}_{k-j} = \hat{a}_k$$
$$k = 1, 2, \cdots, N \tag{5.1.6}$$

其中 $\hat{w}_{k-j}$ 和 $\hat{a}_k$ 都是估计值的简写，其含意见第四章，即 $\hat{w}_{k-j} = E_\beta(w_{k-j} | \boldsymbol{w}_{N-d}), \hat{a}_k = E_\beta(a_k | \boldsymbol{w}_{N-d})$, $\boldsymbol{\beta}$ 是 (5.1.2) 式的参数向量. 至于 $\hat{a}_k$ 的具体计算方法,在第四章已经介绍过了(见第四章 §2 第二小节). 当 N 较大时，这样给出的 a_k 与真实的样本值很接近,在进行检验时，可以用 $a_1, a_2, \cdots, a_N$, 也可以用 $a_{Q+1}, a_{Q+2}, \cdots$, a_N.

§2. 自相关函数检验法

设 $a_1, a_2, \cdots, a_n$ 是白噪声序列的一段样本值,令

$$\hat{\gamma}_k(n, a) = \frac{1}{n} \sum_{j=1}^{n-k} a_j a_{j+k}$$

$$\hat{\rho}_k(n, a) = \frac{\hat{\gamma}_k(n, a)}{\hat{\gamma}_0(n, a)} \tag{5.2.1}$$

按照第三章 §1 的介绍，$(\sqrt{n}\hat{\rho}_1(n, a), \sqrt{n}\hat{\rho}_2(n, a), \cdots, \sqrt{n}\hat{\rho}_K(n, a))$ 的分布渐近于正态 $\mathcal{N}(0, I_K)$. 所以当 n 较大时，这 K 个量近似为相互独立的正态 $\mathcal{N}(0, 1)$ 的随机变量. 于是检验 $a_1, a_2, \cdots, a_n$ 是否为白噪声的样本值问题，便转化成检验上述 K 个量是否为 K 个独立正态 $\mathcal{N}(0, 1)$ 分布的随机变量的问题. 令

$$Q_K = \sum_{k=1}^{K} [\sqrt{n}\,\hat{\rho}_k(n, a)]^2 = n\sum_{k=1}^{K} \hat{\rho}_k^2(n, a) \qquad (5.2.2)$$

则它又化成检验 Q_K 是否是自由度为K的中心 χ^2 分布的问题．这样，便可以用 χ^2_K 表进行检验了(见[18])．现在具体叙述如下．以"a_t 为白噪声"做原假设，以 α 为显著性水平，取半轴 $(\chi^2_{K,\alpha}, \infty)$ 为否定域，根据 α 和K 由统计表查出相应的 $\chi^2_{K,\alpha}$ 之值．然后计算出 Q_K 之值．若 $Q_K^2 \leqslant \chi^2_{K,\alpha}$，则肯定原假设，即认为模型(5.1.1)与被考查的序列 z_t 拟合较好．若 $Q_K > \chi^2_{K,\alpha}$，则否定原假设，即认为拟合得不好．

现在将常用的 $\chi^2_{K,\alpha}$ 之值列于表 5.2.1．在实际应用中，当 n 在数百以上时，K 可取20～30，显著性水平 α 常用 0.05 和 0.01．

表 5.2.1 $\chi^2_{K,\alpha}$ 表

$\chi^2_{K,\alpha}$ \ K / α	20	21	22	23	24	25	26	27	28	29	30
0.05	31.4	32.7	33.9	35.2	36.4	37.7	38.9	40.1	41.3	42.6	43.8
0.01	37.6	38.9	40.3	41.6	43.0	44.3	45.6	47.0	48.3	49.6	50.9

在算出 $\hat{\rho}_k(n, a)$ 的基础上，也可以用一种比 χ^2 检验较粗的方法进行检验．由于$\sqrt{n}\,\hat{\rho}_k(n, a)\ (k=1, 2, \cdots, K)$ 近似为独立正态 $\mathscr{N}(0, 1)$ 分布，可以考察一下这K个值，其中绝对值不大于 1 的比例是否约为 68.3%，不大于 2 的是否为 95.5%；因为正态变量落在 $(-\sigma, \sigma)$ 与 $(-2\sigma, 2\sigma)$ 之间的概率分别是 68.3% 与 95.5%．若经检查后，上述两个比例在相应概率值附近，可认为估计模型拟合得较好，否则就认为较差．仍以(2.1.16 a)为例，据 300 个数据的模型精估计的结果是 $w_t = a_t - 0.42\,a_{t-1}$．用这一模型对 a_t 进行估计，得到 300 个"残量"值，再由这 300 个数据计算出样本相关函数值，我们把前 40 个值列在表 5.2.2 里．从这些数据中可以查出，有 28/40=70.0% 的 $\hat{\rho}_k(300, a)$ 不超过 $\frac{1}{\sqrt{300}}=0.058$，这与 68.3% 是相近的；有 39/40=97.5% 的 $\hat{\rho}_k$

(300, a) 不超过 $2\times0.058=0.116$，这与95.5%也相近．因此经这种粗略的检验认为可以接受第四章所给出的（2.1.16 a）估计模型．

表5.2.2　(2.1.16a)序列的模型拟合的残量样本自相关函数

k	1	2	3	4	5	6	7	8
$\hat{\rho}_k(300, a)$	0.083	0.015	0.064	−0.083	0.056	0.052	−0.040	−0.007
k	9	10	11	12	13	14	15	16
$\hat{\rho}_k(300, a)$	−0.089	−0.041	0.058	−0.012	0.035	−0.047	0.031	0.054
k	17	18	19	20	21	22	23	24
$\hat{\rho}_k(300, a)$	0.005	0.137	0.095	0.091	−0.016	−0.073	0.078	0.021
k	25	26	27	28	29	30	31	32
$\hat{\rho}_k(300, a)$	−0.024	−0.028	−0.053	−0.104	0.054	0.023	0.082	0.056
k	33	34	35	36	37	38	39	40
$\hat{\rho}_k(300, a)$	−0.013	0.015	0.055	−0.003	−0.001	0.022	−0.051	−0.056

若采用 χ^2 检验，我们算出

$$Q_{20}=300\times\sum_{k=1}^{20}\hat{\rho}_k^2(300, a)=24.91$$

以水平 $\alpha=0.05$ 查 $\chi^2_{K,\alpha}$ 表知 $P(Q_{20}>31.4)=0.05$．因 $24.91<31.4$，故可以接受原假设，即认为拟合得好．

§3. 周期图检验法

设 w_t 为一均值为零的随机序列，令 $0\leqslant\lambda\leqslant\frac{1}{2}$，

$$\begin{cases} A_n(\lambda) = \sqrt{\dfrac{2}{n}} \sum\limits_{k=1}^{n} w_k \cos 2\pi\lambda k \\ B_n(\lambda) = \sqrt{\dfrac{2}{n}} \sum\limits_{k=1}^{n} w_k \sin 2\pi\lambda k \end{cases} \tag{5.3.1}$$

当取 $\lambda = \frac{1}{n}, \frac{2}{n}, \cdots, \left[\frac{n}{2}\right] \Big/ n$ $\left(\left[\frac{n}{2}\right]\right.$表示 $\frac{n}{2}$ 的整数部分$\left.\right)$ 时，有

$$\frac{2}{n} \sum_{k=1}^{n} (\cos 2\pi\lambda k)^2 = 1, \quad \frac{2}{n} \sum_{k=1}^{n} (\sin 2\pi\lambda k)^2 = 1 \tag{5.3.2}$$

又令

$$I_n(\lambda) = A_n^2(\lambda) + B_n^2(\lambda) = \left| \sqrt{\frac{2}{n}} \sum_{k=1}^{n} w_k e^{i2\pi\lambda k} \right|^2 \tag{5.3.3}$$

$I_n(\lambda)$ 称为 w_t 的周期图，它又可以写成

$$\begin{aligned} I_n(\lambda) &= \frac{2}{n} \sum_{k=1}^{n} w_k^2 + \frac{4}{n} \sum_{k=1}^{n-1} \left(\sum_{j=1}^{n-k} w_j w_{k+j} \right) \cos 2\pi\lambda k \\ &= 2 \sum_{k=1-n}^{n-1} \hat{\gamma}_k \cos 2\pi\lambda k \end{aligned} \tag{5.3.4}$$

这里的 $\hat{\gamma}_k = \hat{\gamma}_{-k}$ 是 $w_1, w_2, \cdots, w_k$ 的样本自协方差函数. 若 $\sum\limits_{k=-\infty}^{\infty} |\gamma_k| < \infty$，可在 (5.3.4) 式两边求均值得

$$EI_n(\lambda) = 2 \sum_{k=1-n}^{n-1} \left(1 - \frac{|k|}{n}\right) \gamma_k \cos 2\pi\lambda k \tag{5.3.5}$$

记

$$g_H(\lambda) = 2 \sum_{k=-\infty}^{\infty} \gamma_k \cos 2\pi\lambda k = 2\left(\gamma_0 + 2 \sum_{k=1}^{\infty} \gamma_k \cos 2\pi\lambda k\right) \tag{5.3.6}$$

由于此处 λ 限定在 $\left[0, \frac{1}{2}\right]$ 上，所以称 $g_H(\lambda)$ 为半功率谱密度. 从以上两式可知

$$\lim_{n\to\infty} EI_n(\lambda) = g_H(\lambda) \tag{5.3.7}$$

这说明周期图是半功率谱密度的渐近无偏估计. 又由于

$$\int_0^{\frac{1}{2}} \cos 2\pi\lambda k d\lambda = 0 \qquad (k > 0)$$

以及 $\sum_{k=-\infty}^{\infty} |\gamma_k| < \infty$，对 (5.3.6) 式进行积分可得

$$\int_0^{\frac{1}{2}} g_H(\lambda) d\lambda = \gamma_0 \tag{5.3.8}$$

特别，当 w_t 为白噪声时，$\gamma_k = 0$ $(k > 0)$，所以

$$g_H(\lambda) = 2\gamma_0 \tag{5.3.9}$$

正如随机变量的概率密度函数 $f(x)$（假定它存在）与其分布函数 $F(x)$ 的关系一样，半功率谱分布函数被定义为

$$G_H(\lambda) = \int_0^{\lambda} g_H(\mu) d\mu \qquad \left(0 \leqslant \lambda \leqslant \frac{1}{2}\right) \tag{5.3.10}$$

显然 $G_H\left(\frac{1}{2}\right) = \gamma_0$.

根据上述 (5.3.7) 式 $I_n(\lambda)$ 的均值收敛于 $g_H(\lambda)$，似乎可用 $I_n(\lambda)$ 做为 $g_H(\lambda)$ 的估计，遗憾的是这个估计不是相容的. 以正态白噪声 a_t 为例，当 $n \to \infty$ 时，利用 Bartlett 公式（即 (3.1.7) 式或 (3.1.8) 式以及附录 §2 的 (2.4) 式）可知 $\tilde{I}_n(\lambda) = I_n(\lambda) - 2\gamma_0$ 的方差满足

$$\lim_{n\to\infty} \operatorname{var} \tilde{I}_n(\lambda) = \lim_{n\to\infty} E\left(2 \sum_{k=1-n}^{n-1} \tilde{\gamma}_k \cos 2\pi\lambda k\right)^2$$
$$= (2\gamma_0)^2 = 4\sigma_a^2$$

此式说明无论怎样加大样本长度 n，估计误差 $\tilde{I}_n(\lambda)$ 的方差开根量与被估计量 $g_H(\lambda) = 2\gamma_0$ 本身几乎一样大. 因此，这种估计是很不理想的.

现在令

$$\mathscr{I}_n(\lambda) = \int_0^{\lambda} I_n(\mu) d\mu \qquad \left(0 \leqslant \lambda \leqslant \frac{1}{2}\right) \tag{5.3.11}$$

称之为序列 w_t 的积分周期图. 关于积分周期图，在上面相同的条件下，已被证明（参看 [5]）它是半功率谱分布函数 $G_H(\lambda)$ 的相容估计，而且给出了最大误差随机变量

$$\max_{0\leqslant\lambda\leqslant\frac{1}{2}} \frac{\sqrt{n}}{2}|\mathscr{I}_n(\lambda)-G_H(\lambda)|$$

的渐近分布. 下面我们仅叙述在这里要用到的有关的一些结果.

设平稳序列是正态白噪声，不妨设 $Ea_t^2=1$，于是 $G_H(\lambda)=2\lambda$. 当 $n\to\infty$ 时，

$$\max_{0\leqslant\lambda\leqslant\frac{1}{2}} \frac{\sqrt{n}}{2}|\mathscr{I}_n(\lambda)-2\lambda|=\max_{0\leqslant\lambda\leqslant\frac{1}{2}} \sqrt{n}\left|\frac{1}{2}\mathscr{I}_n(\lambda)-\lambda\right| \tag{5.3.12}$$

的分布收敛于随机变量

$$\max_{0\leqslant\lambda\leqslant\frac{1}{2}} |\zeta(\lambda)| \tag{5.3.13}$$

的分布，其中 $\zeta(\lambda)$ 为 Brown 运动，它是一种特殊的正态过程；即满足条件

$$\begin{cases} E\zeta(\lambda)=0 & \lambda\geqslant 0 \\ E\zeta(\lambda)\zeta(\mu)=\min(\lambda,\mu) & \lambda,\mu\geqslant 0 \end{cases} \tag{5.3.14}$$

$\max\limits_{0\leqslant\lambda\leqslant\frac{1}{2}} |\zeta(\lambda)|$ 的分布为

$$\begin{aligned} \Delta_0(x) &= P\left(\max_{0\leqslant\lambda\leqslant\frac{1}{2}} |\zeta(\lambda)|\leqslant x\right) \\ &= \sum_{k=-\infty}^{\infty} (-1)^k[\Phi(\sqrt{2}(2k+1)x) \\ &\quad -\Phi(\sqrt{2}(2k-1)x)] \end{aligned} \tag{5.3.15}$$

其中 $\Phi(y)$ 是正态 $\mathscr{N}(0,1)$ 的分布函数，即

$$\Phi(y)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y} e^{-\frac{x^2}{2}}dx$$

我们在这里举出几个 $\Delta_0(x)$ 的取值

表 5.3.1　[0,1/2] 上 Brown 运动最大模的分布函数值

$\Delta_0(x)$	0.999	0.99	0.95	0.90	0.75
x	2.46	1.99	1.58	1.38	1.08

为了检验 a_t 是否是白噪声列（而且 $Ea_t^2=1$），先用 a_t 的样

本值 $a_1, a_2, \cdots, a_n$ 计算出 (5.3.12) 式的取值，若选定 $\alpha = 1 - \Delta_0(x)$ 为显著性水平，就可以作假设检验了．比如取 $\alpha = 0.01$，则 $\Delta_0(x) = 0.99$，由表5.3.1查出 $\Delta_0(1.99) = 0.99$，若由 (5.3.12) 式求出的样本值比 1.99 小，就承认 a_t 为白噪声，且 $Ea_t^2 = 1$，否则就不承认 a_t 为白噪声．

在实际应用这种方法时，还会遇到以下几个问题．

首先是关于 $\mathscr{I}_n(\lambda)$ 的积分值计算问题．依照 $\mathscr{I}_n(\lambda)$ 的定义公式 (5.3.11) 式，须求出每个 $I_n(\mu)\left(\mu \in \left[0, \frac{1}{2}\right]\right)$ 的值．在实际应用时，只能用数值积分代替，最简单的办法是用等间隔分点的 Riemann 和代替，即令 $\mathscr{I}_n^{(m)}(0) = 0$ 及

$$\mathscr{I}_n^{(m)}(\lambda) = \frac{1}{m}\sum_{k=1}^{[m\lambda]} I_n\left(\frac{k}{m}\right), \qquad 0 \leqslant \lambda \leqslant \frac{1}{2} \quad (5.3.16)$$

其中 $[m\lambda]$ 表示 $m\lambda$ 的整数部分，m 为适当大的正整数（即分点的个数）．我们称 $\mathscr{I}_n^{(m)}(\lambda)$ 为序列的积累周期图，它是阶梯函数，即满足

$$\mathscr{I}_n^{(m)}(\lambda) = \mathscr{I}_n^{(m)}\left(\frac{k}{m}\right), \qquad \frac{k}{m} \leqslant \lambda < \frac{k+1}{m}$$

当 $m \to \infty$ 时，由积分定义可知

$$\lim_{m\to\infty} \mathscr{I}_n^{(m)}(\lambda) = \mathscr{I}_n(\lambda)$$

因此，在进行检验时，可以用积累周期图 $\mathscr{I}_n^{(m)}(\lambda)$ 近似积分周期图 $\mathscr{I}_n(\lambda)$．当样本长度 n 很大时，又常取 $m = n$，即

$$\mathscr{I}_n^{(n)}(\lambda) = \frac{1}{n}\sum_{k=1}^{[n\lambda]} I_n\left(\frac{k}{n}\right), \qquad 0 < \lambda \leqslant \frac{1}{2} \quad (5.3.17)$$

而 $\mathscr{I}_n^{(n)}(0) = 0$．

其次，在上述的检验中都把 $Ea_t^2 = 1$ 列入原假设，若 $Ea_t^2 = \sigma_a^2 \neq 1$ 又应当怎样检验呢？这时，我们要做一个变换，即令 $b_t = \frac{a_t}{\sigma_a}$，则 $Eb_t^2 = 1$，而 b_t 仍是白噪声列，再用以上公式和方法检验 $b_1, b_2, \cdots, b_n\left(b_k = \frac{a_k}{\sigma_a}\right)$ 是否是正态白噪声列（而且 $Eb_t^2 = 1$）就

可以了.

但是，在实际应用中，一般 σ_a^2 是我们待估计的参数，并不知道其真值. 这时又不得不以 $\hat{\sigma}_a^2$ 代替 σ_a^2 使用. 由于 $\hat{\sigma}_a^2$ 是 σ_a^2 的相容估计，所以当 n 较大时，$\hat{\sigma}_a^2$ 与 σ_a^2 相当接近，真正的检验是对 $\frac{a_1}{\hat{\sigma}_a}$, $\frac{a_2}{\hat{\sigma}_a}$, $\cdots$, $\frac{a_n}{\hat{\sigma}_a}$ 进行的.

最后，我们归纳一下用周期图法检验 a_t 序列是否是白噪声的步骤.

1. 利用已给的估计模型和量测数据，求出样本值 $a_1, a_2, \cdots, a_n$. 以"a_t 为正态白噪声，$Ea_t=0$, $Ea_t^2=\hat{\sigma}_a^2$"为原假设，以"原假设不成立"为对立假设.

2. 由(5.3.16)式(或(5.3.17)式)计算出 $\frac{1}{2\hat{\sigma}_a^2}\mathscr{I}_n^{(m)}\left(\frac{1}{m}\right)$, $\frac{1}{2\hat{\sigma}_a^2}\mathscr{I}_n^{(m)}\left(\frac{2}{m}\right)$, $\cdots$, $\frac{1}{2\hat{\sigma}_a^2}\mathscr{I}_n^{(m)}\left(\frac{[m/2]}{m}\right)$.

3. 适当地选择显著性水平 α，则 $\Delta_0(x)=1-\alpha$，由 $\Delta_0(x)$ 的取值表(表5.3.1)查出相应于 α 的值 x_α. 在平面坐标纸上绘出 $(\lambda, \eta(\lambda))$ 坐标轴和对角线，并绘出以对角线为中心线的带形区域

$$\lambda-\frac{x_\alpha}{\sqrt{n}}\leqslant\eta(\lambda)\leqslant\lambda+\frac{x_\alpha}{\sqrt{n}}$$

(见图 5.3.1). 若二维点列 $\left(\frac{k}{m}, \frac{1}{2\hat{\sigma}_a^2}\mathscr{I}_n^{(m)}\left(\frac{k}{m}\right)\right)$ $\left(1\leqslant k\leqslant\left[\frac{m}{2}\right]\right)$ 都落在这条带子内部，就接受原假设，否则就拒绝原假设，也就是认为所给的估计模型与量测数据序列拟合得不好.

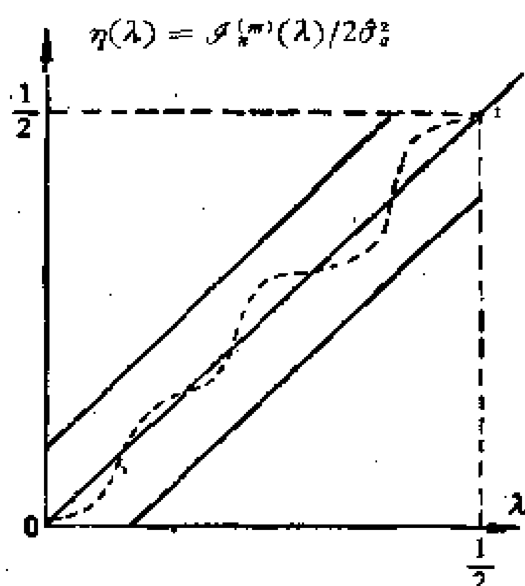


图 5.3.1

根据前面的阐述，若原假设是正确的，则应有近似公式

$$P\left(\max_{1\leqslant k\leqslant\left[\frac{m}{2}\right]}\left|\frac{1}{2\hat{\sigma}_a^2}\cdot\mathscr{I}_n^{(m)}\left(\frac{k}{m}\right)-\frac{k}{m}\right|\leqslant\frac{x_\alpha}{\sqrt{n}}\right)\cong\Delta_0(x_\alpha)$$

因此，当本来模型正确时（或者说拟合得好），由于它们是随机量使我们做出错误的判断（即认为拟合得不好），这类事件发生的概率近似为 α.

§4. 估计模型的一种改进方法

如果经过前面所述的检验判断模型拟合得不好，那么一般说来，使用这个模型去解决实际中的预报、控制等问题时，其效果也不会好. 为此，我们应设法改进原来假想的模型，或用其他方法给出对随机序列的进一步描述.

假设 z_t 的设想模型是 ARIMA(p, d, q) 型，即 $w_t=\nabla^d z_t$ 是 ARMA(p, q) 序列. 为了讨论得更广泛些，我们假定 z_t 的模型是

$$\varphi(B)\nabla^d z_t=c+\theta(B)a_t \tag{5.4.1}$$

其中若 $d=0$，则 $c=(1-\varphi_1-\varphi_2-\cdots-\varphi_p)\mu$，若 $d>0$ 则 $c=0$. 运用上节的方法进行检验，若认为假想模型拟合得不好，则可利用由假想模型和量测数据 $z_1, z_2, \cdots, z_N$ 计算出的“残量”值 $a_1, a_2, \cdots, a_N$，再作一次识别和估计. 比如得到的结果是 a_t 为 ARIMA(P, D, Q)序列，即

$$\Phi(B)\nabla^D a_t=\Theta(B)b_t \tag{5.4.2}$$

这里用大写字母 P, D, Q, Φ, Θ 及 $b_t(Eb_t^2=\sigma_b^2)$ 等，是为了区别于原有的模型，它们的含意仍如第二章所述. 也就是说，我们还可以用(5.4.2)式的模型来拟合“残量”序列. 当然，同样也可以对这个模型按前一节的方法进行检验，若被接受，就得到了一个新的改进后的估计模型为

$$\varphi(B)\Phi(B)\nabla^{d+D}z_t=\Phi(B)\nabla^D c+\theta(B)\Theta(B)b_t \tag{5.4.3}$$

其中若 $D>0$，则 $\nabla^D c=0$，否则 $\Phi(B)c=(1-\Phi_1-\Phi_2-\cdots-\Phi_P)c$.

从原则上说，如果 (5.4.2) 式的模型仍不被接受，还可以继续上述的改进步骤. 但在实际应用中，很可能实际序列已不能用这

类模型来拟合了，需要对序列作其他方面的分析.

在第三章中介绍模型初步识别的方法时，混合模型阶数(p,q)的确定，采取从低阶到高阶逐步尝试的办法. 每尝试一种模型，都要进行模型检验，如果不合要求，就要提高阶数. 显然本节的方法对于这类模型识别是有实际意义的.

§5. 分段拟合与适时拟合

对于一个从实际问题中提出的时间序列，我们总希望能用单一的 ARIMA 序列来描述它. 或者说，至少在我们所关心的有限时间区间内，能用一种 ARIMA 序列来拟合它，这样会给应用带来很大的方便. 但就很多真实的时间序列而言，把它的某一小段用单一的 ARIMA 序列来拟合是容许的，而若要求在更长的区间里用同一模型来拟合，就很难通过检验而被接受. 在这种情况下，虽说按上一节的方法不断改进，有时也可能通过检验被接受，但是，用阶数过高的模型来拟合不太长的样本值序列，并不是解决问题的好办法. 因为这样会大大降低每个参数的估计精度. 如果企图用继续加大样本长度来弥补这个损失，在很多实际情况中也不理想，因为，真实时间序列的结构十分复杂，随着所考虑的时间区间的增加，原有的阶数又可能不够用了. 为了使这种比较复杂的(一般不是平稳的)时间序列也能用 ARIMA 序列近似描述，我们提供以下的分段拟合与适时拟合的基本想法.

一、分段拟合方法

对某些难于用单一的 ARIMA 序列描述的真实时间序列，可以进行分段拟合，也就是说，把数据分成若干段，在每段内用一种 ARIMA 序列拟合，在不同段内可以出现不同的模型. 在确定了分段方案以后，各段内的识别与估计完全按照前面已有的方法进行. 本节所要介绍的是如何确定分段方案. 实际上，我们只须讨论在对一个长度为N的数据列进行拟合时，怎样判断有无必要分成两段模型的问题. 至于更多的分段问题，可以由此推广而得到解

决. 这种判断和分段拟合方法，对于随机序列的模拟和随机信号变异的检测也有参考价值.

设有一个长度为N的样本数据列，我们可先将数据大致平分为两段（或者按其物理根据给出分段位置），即 $N_1+N_2=N$，$N_1\cong N_2$，对各段数据分别进行时序分析. 然后利用这些分析结果(即各自的 ARIMA 模型)来判断这种分段处理有无必要.

关于这一问题，我们只讨论一种特殊情况，即两段拟合模型的阶数是相同的，只是它们各自的参数可能不同. 假定第一段的真实参数与估计参数分别用 $\boldsymbol{\alpha}(1)=(\boldsymbol{\beta}(1)^\tau,\sigma_a^2(1))^\tau$ 与 $\hat{\boldsymbol{\alpha}}(1)=(\hat{\boldsymbol{\beta}}^\tau(1),\hat{\sigma}_a^2(1))^\tau$ 表示；第二段分别用 $\boldsymbol{\alpha}(2)$, $\hat{\boldsymbol{\alpha}}(2)$ 表示. 根据估计的渐近正态分布性质可知

$$\hat{\alpha}_j(1)\sim\mathcal{N}(\alpha_j(1),\sigma_j^2(1))\quad \hat{\alpha}_j(2)\sim\mathcal{N}(\alpha_j(2),\sigma_j^2(2))$$

其中 $\alpha_j(i)$, $\hat{\alpha}_j(i)$ 分别是 $\boldsymbol{\alpha}(i)$ 与 $\hat{\boldsymbol{\alpha}}(i)$ 的第 j 个分量，$\sigma_j^2(i)$ 是信息阵的第 j 行 j 列之元素. 在此我们还假定 $\hat{\boldsymbol{\alpha}}(1)$ 与$\hat{\boldsymbol{\alpha}}(2)$ 是渐近独立的. 这样可以借助于假设检验方法，判断 $\boldsymbol{\alpha}(1)$ 与 $\boldsymbol{\alpha}(2)$ 是否不相同. 以"参数 $\boldsymbol{\alpha}(1)=\boldsymbol{\alpha}(2)$"为原假设，以"参数 $\boldsymbol{\alpha}(1)\neq\boldsymbol{\alpha}(2)$"为对立假设，根据独立变量和的分布可知，若原假设成立，则

$$\alpha_j(1)-\alpha_j(2)\sim\mathcal{N}(0,\sigma_j^2(1)+\sigma_j^2(2))$$

因而$|\alpha_j(1)-\alpha_j(2)|$不超过 $2\sqrt{\sigma_j^2(1)+\sigma_j^2(2)}$ 的概率为 95.5%，而超过的概率只是 4.5%. 在信息矩阵公式中找出 $\sigma_j^2(i)=\mathrm{var}(\hat{\alpha}_j(i))$ 的表达式，并以估计量 $\hat{\boldsymbol{\alpha}}(i)$ 代替公式中的真参数，然后检查 $|\hat{\alpha}_j(1)-\hat{\alpha}_j(2)|$ 是否超过 $2\sqrt{\hat{\sigma}_j^2(1)+\hat{\sigma}_j^2(2)}$，若没有超过，则认为原假设成立，否则拒绝原假设，也就是说，认为有必要进行分段拟合.

例如，某时间序列 z_t 的样本长度为 368，经过识别后为 AR(1, 0) 序列，其估计模型为

$$z_t+0.09z_{t-1}=a_t$$

参数 $\hat{\varphi}_1=-0.09$. 但是，经过前一节所述的模型检验后，认为拟合得不好. 为了改进拟合效果，我们可以采用分段拟合. 把 368 个数据分成两段，每段 184 个数据. 分别进行参数估计，其结果是

$\phi_1(1) = -0.29$，$\phi_1(2) = -0.03$．根据第四章§4的公式可知

$$\mathrm{var}(\hat{\varphi}_1) \cong \frac{1}{N}(1-\varphi_1^2)$$

以 $N=184$ 以及 $\phi_1(1)=-0.29$ 和 $\phi_1(2)=-0.03$ 分别代人上式可得

$$\sigma_a^2(1) \cong \frac{1}{184}[1-(0.29)^2] \cong \frac{1}{202}$$

$$\sigma_a^2(2) \cong \frac{1}{184}[1-(0.03)^2] \cong \frac{1}{184}$$

$$[\sigma_a^2(1)+\sigma_a^2(2)]^{\frac{1}{2}} \cong \sqrt{0.0104} \cong 0.102$$

但是 $|\varphi(1)-\varphi(2)| = 0.29-0.03=0.26$，超过 $2\times 0.102 = 0.204$．因此我们认为有必要将368个数据分段进行参数估计．

二、适时拟合方法

如果经过前面的检验判断 z_t 不适合用单一的 ARIMA 模型描述，而描述的目的又在于解决适时最佳预报或控制等问题时，分段拟合往往是徒劳的．这时我们只关心靠近现在的一段历史和不远的未来时刻 z_t 之间的联系，即使进行分段拟合，也只须要后一段的拟合结果．根据类似于滤波理论中限定记忆滤波的方法（见[20]），我们提出适时拟合的方法．也就是说，为了掌握不远的未来时刻的序列值与靠近现在的一段历史值的联系，我们总是根据到目前 t 时刻为止的 M 个样本值 $z_{t-M+1}, z_{t-M+2}, \cdots, z_t$ 在线（即适时）地进行 z_t 的模型参数估计，不断为预报或控制提供新的模型参数．

例如为了报时，用几架天文仪器对同一个天文目标进行持续观测．各架仪器都有观测误差，为了提高观测精度，常常把这几架仪器观测的结果进行加权平均（如算术平均）确定出"真值"，而把每架仪器与这"真值"之差称为该仪器的观测误差，以 δ_{ti} 表示，这种误差叫做人仪差，其中 t 表示第 t 次（例如表示第 t 个月），i 表示第 i 架．如果根据N个月的 δ 历史值，即 $\delta_{t-1,i}, \cdots, \delta_{t-N,i}$ 给 δ_{ti} 一个预报值 $\hat{\delta}_{ti}$，那么把这误差值加到观测值中去再去参加

平均从而得到第 t 个月的"真值",会提高精度. 因此,可用时间序列的方法建立 δ_{ti} 的模型并预报它. 为方便,下面都记 δ_{ti} 为 δ_t.

我国天文工作者曾用序列的 12 个算术平均值作为预报值,即

$$\hat{\delta}_t = \frac{1}{12}\sum_{j=1}^{12}\delta_{t-j}$$

也就是说,把 δ_t 看成均值不为零的白噪声.

现在用本书的方法来做. 对数据的分析结果表明,如果限制数据长度不超过 48,几乎所有的样本 $\delta_{t-48}, \delta_{t-47}, \cdots, \delta_{t-1}$ 都可用 AR(1, 0) 或 AR(2, 0) 模型来很好地拟合;但是对不同的 t 参数的变化较大. 根据这种分析,我们用适时拟合的方法. 这就是在时刻 t,用 $\delta_{t-48}, \cdots, \delta_{t-1}$ 的样本值求出 AR(1, 0) 或 AR(2, 0) 的参数估计 $\hat{\boldsymbol{\varphi}}(t)$ 来,以此为参数按第七章中的方法预报 δ_t.

例如,定模型为 AR(1, 0),则它为

$$(\delta_t - \mu(t)) - \hat{\varphi}(t)(\delta_{t-1} - \mu(t)) = a_t$$

或

$$\delta_t - \hat{\varphi}(t)\delta_{t-1} - \mu(t)(1 - \hat{\varphi}(t)) = a_t$$

其中

$$\mu(t) = \frac{1}{48}\sum_{j=1}^{48}\delta_{t-j}$$

$$\hat{\varphi}(t) = \sum_{j=1}^{47}(\delta_{t-j} - \mu(t))(\delta_{t-j+1} - \mu(t)) \Big/ \sum_{j=1}^{48}(\delta_{t-j} - \mu(t))^2$$

而预报值为

$$\hat{\delta}_t = \mu(t) + \hat{\varphi}(t)(\delta_{t-1} - \mu(t))$$

在时刻 $t+1$,我们去掉最早的数据 δ_{t-48},添上新测得的数据 δ_t,然后重新估计一次模型参数,用这模型来预报 δ_{t+1},以后如此继续下去.

根据上海天文台的计算结果表明,用这种方法比用算术平均预报方法效果要好.

举一数值实例. 对一台等高仪的数据进行处理,将 1963 年至 1975 年间的历史数据用算术平均法进行预报,算得预报误差均方根值为 2.6 毫秒. 采用上述方法,若以 AR(1, 0)拟合,其预报误差的均方根值为 2.1 毫秒,用 AR(2,0) 为 2.0 毫秒. 如果将数据按半个月一次处理,甚至十天一次处理,还可进一步提高预报精度.

第六章　序列的合成、分解及季节性模型

§1. 序列的合成

在实际问题中，常常遇到两个以上的相互独立的随机序列相叠加的问题．例如信号接收机的输出量，就是由有用信号与干扰噪声叠加而成的．一般干扰噪声又是几种相互独立的噪声源所产生的噪声量的叠加，比如接收机内部的热噪声，信号在传输中受到干扰产生的噪声等等．以这样的实际问题为背景，我们提出以下的时间序列分析中的合成问题．

我们只考虑两个相互独立 ARIMA 序列的叠加问题，多个的情况是这种情况的简单推广．假定

$$z_t = z_t^{(1)} + z_t^{(2)} \tag{6.1.1}$$

其中 $z_t^{(1)}$ 与 $z_t^{(2)}$ 是两个相互独立的 ARIMA 序列．我们要讨论 z_t 是怎样的序列，它的模型阶数及参数值与两个分序列 $z_t^{(i)}(i=1,2)$ 的模型及参数的关系．

先考虑 $z_t^{(1)}$ 与 $z_t^{(2)}$ 都是滑动平均序列的情况．设

$$z_t^{(i)} = \Theta^{(i)}(B)a_t^{(i)} \tag{6.1.2}$$

其中 $\Theta^{(i)}(B) = 1 - \Theta_1^{(i)}B - \cdots - \Theta_{q_i}^{(i)}B^{q_i}\quad(i=1,2)$，令

$$q = \max\{q_1, q_2\}$$

以及 $\gamma_k^{(i)} = E(z_t^{(i)}z_{t+k}^{(i)})$．由于 $z_t^{(1)}$ 与 $z_t^{(2)}$ 相互独立，所以

$$\gamma_k = E(z_tz_{t+k}) = E(z_t^{(1)}z_{t+k}^{(1)}) + E(z_t^{(2)}z_{t+k}^{(2)}) = \gamma_k^{(1)} + \gamma_k^{(2)} \tag{6.1.3}$$

又由于 $\gamma_k^{(1)}$，$\gamma_k^{(2)}$ 都是截尾的，所以 γ_k 也是截尾的，而且显然当 $k>q$ 时 $\gamma_k=0$．因此，z_t 是 MA$(0,q)$ 型序列，暂时记做

$$z_t = \theta(B)a_t \tag{6.1.4}$$

根据附录 §1 可知，只要 $\Theta^{(i)}(B)$ 中有一个是可逆的，$\theta(B)$ 就一定

是可逆的. $\theta(B)$ 的系数可由 $\Theta^{(i)}(B)(i=1,2)$ 的系数定出. 事实上,因为已知 (6.1.1) 式的 z_t 可表为(6.1.4)式的模型,所以它们的自协方差函数应该相等,这样便可建立如下的关系式,即

$$\begin{aligned}\sigma_a^2(1+\theta_1^2+\cdots+\theta_q^2) &= \sigma_1^2[1+(\Theta_1^{(1)})^2+\cdots+(\Theta_{q_1}^{(1)})^2] \\ &\quad + \sigma_2^2[1+(\Theta_1^{(2)})^2+\cdots+(\Theta_{q_2}^{(2)})^2] \\ \sigma_a^2(-\theta_k+\theta_1\theta_{k+1}+\cdots+\theta_{q-k}\theta_q) &= \sigma_1^2(-\Theta_k^{(1)}+\Theta_1^{(1)}\Theta_{k+1}^{(1)}+\cdots \\ &\quad +\Theta_{q_1-k}^{(1)}\Theta_{q_1}^{(1)}) + \sigma_2^2(-\Theta_k^{(2)} \\ &\quad + \Theta_1^{(2)}\Theta_{k+1}^{(2)}+\cdots+ \\ &\quad \Theta_{q_2-k}^{(2)}\Theta_{q_2}^{(2)}),\ k=1,2,\cdots,q\end{aligned} \tag{6.1.5}$$

其中 $\sigma_i^2=E(a_t^{(i)})^2$. 这是 $q+1$ 阶非线性联立方程式, 右边都是已知数, 左边有 $q+1$ 个未知量. 一般说来, 它的解可以有多个, 我们只能从其中挑出满足可逆条件的解. 在 $q=1$ 时, 容易给出 σ_a^2 和 θ_1 的解的表达式. 除此以外, 通常求解 (6.1.5) 式都要用数值解法. 一般情况下 $\theta_q \neq 0$; 但在 $q_1=q_2$ 而且 $\Theta_q^{(1)}\sigma_1^2=-\Theta_q^{(2)}\sigma_2^2$ 时 $\theta_q=0$, 即(6.1.4) 式是退化的 q 阶滑动平均序列.

现在考虑分序列为一般 ARIMA 序列的情况. 设 $z_t^{(i)}$ 为 ARIMA (p_i, d_i, q_i) 序列,其参数为 $\beta^{(i)}(i=1,2)$, 即

$$\varphi^{(i)}(B)\nabla^{d_i}z_t^{(i)}=\theta^{(i)}(B)a_t^{(i)} \tag{6.1.6}$$

并且假定 $\varphi^{(1)}(B)$ 与 $\varphi^{(2)}(B)$ 没有相同的根. 在有相同根的情形, 读者阅下文后自能处理. 令 $d=\max\{d_1, d_2\}$, 则

$$\begin{aligned}\varphi^{(1)}(B)\varphi^{(2)}(B)\nabla^d z_t &= \varphi^{(1)}(B)\varphi^{(2)}(B)\nabla^d z_t^{(1)} \\ &\quad + \varphi^{(1)}(B)\varphi^{(2)}(B)\nabla^d z_t^{(2)} \\ &= \varphi^{(2)}(B)\theta^{(1)}(B)\nabla^{d-d_1}a_t^{(1)} \\ &\quad + \varphi^{(1)}(B)\theta^{(2)}(B)\nabla^{d-d_2}a_t^{(2)}\end{aligned} \tag{6.1.7}$$

记 $\varphi^{(2)}(B)\theta^{(1)}(B)\nabla^{d-d_1}\equiv\Theta^{(1)}(B)$, $\varphi^{(1)}(B)\theta^{(2)}(B)\nabla^{d-d_2}\equiv\Theta^{(2)}(B)$, 注意 $d-d_1$ 和 $d-d_2$ 至少有一个为零. 于是若令

$$y_t^{(1)}=\Theta^{(1)}(B)a_t^{(1)}, \qquad y_t^{(2)}=\Theta^{(2)}(B)a_t^{(2)} \tag{6.1.8}$$

则其中至少有一个是可逆的. 由前面所述,合成序列

$$y_t=y_t^{(1)}+y_t^{(2)} \tag{6.1.9}$$

为 MA$(q, 0)$ 序列，即

$$y_t = \theta(B)a_t$$

其中

$$q = \max\{p_2 + d - d_1 + q_1, p_1 + d - d_2 + q_2\} \quad (6.1.10)$$

再令

$$\varphi^{(1)}(B)\varphi^{(2)}(B) = \varphi(B), \ p = p_1 + p_2 \quad (6.1.11)$$

可见合成序列 z_t 适合于如下的模型

$$\varphi(B)\nabla^d z_t = \theta(B)a_t \quad (6.1.12)$$

也就是说，两个独立的 ARIMA 序列之和仍是 ARIMA 序列，其阶数和各参数之值由以上诸式给定.

在实际应用中，最常见的合成是 $z_t^{(1)}$ 为 ARIMA(p_1, d_1, q_1) 序列及 $z_t^{(2)}$ 为白噪声，即 $z_t^{(2)} = a_t^{(2)}$. 由上知

$$z_t = z_t^{(1)} + a_t^{(2)}$$

为 ARIMA(p_1, d_1, q) 序列，其中 $q = \max\{q_1, p_1 + d_1\}$.

例如 $\nabla z_t^{(1)} = (1 - \theta^{(1)}B)a_t^{(1)}$, $z_t^{(2)} = a_t^{(2)}$, 则 $z_t = z_t^{(1)} + z_t^{(2)}$ 的模型为

$$\nabla z_t = (1 - \theta B)a_t$$

令 $w_t = \nabla z_t = \nabla z_t^{(1)} + \nabla a_t^{(2)} = a_t^{(1)} - \theta^{(1)}a_{t-1}^{(1)} + a_t^{(2)} - a_{t-1}^{(2)}$，于是由 z_t 的两种等价的模型表达式可得

$$\begin{cases}\gamma_0(w) = (1 + \theta^2)\sigma_a^2 = [1 + (\theta^{(1)})^2]\sigma_1^2 + 2\sigma_2^2 \\ \gamma_1(w) = -\theta\sigma_a^2 = -\theta^{(1)}\sigma_1^2 - \sigma_2^2\end{cases}$$

上式右边都是已知数，暂时用 A 和 B 分别表示，于是上式可简写成

$$\begin{cases}(1+\theta^2)\sigma_a^2 = A \\ -\theta\sigma_a^2 = B\end{cases}$$

消去变元 σ_a^2 可得

$$\theta^2 + \frac{A}{B}\theta + 1 = 0$$

设其解为 x_1 和 x_2，注意 $x_1 x_2 = 1$，不妨设 $|x_1| < 1$，于是解出 $\theta = x_1$, $\sigma_a^2 = -\dfrac{B}{x_1}$，这就是合成序列的模型参数值. 现在给出本

例的一组具体数值：$\theta^{(1)}=0.5$，$\sigma_1^2=\sigma_2^2=1$，则 $A=\frac{B}{4}$，$B=-\frac{3}{2}$. 于是 $x_1=\frac{2}{3}$，$x_2=\frac{3}{2}$，$\theta=\frac{2}{3}$，$\sigma_a^2=\frac{3}{2}\times\frac{3}{2}=\frac{9}{4}$.

§2. 序列的分解

本节讨论一个与序列合成相反的问题，即序列分解问题. 当我们已知几个独立的随机序列的合成序列的模型时，能否求出每个分序列的模型和参数值呢？在实际中这类问题比合成问题遇到得更多. 例如，在水泥生产自动控制中，需要知道窑内某处物料的温度，但是测量时常带有随机误差，如何根据量测序列的统计特性求出真正温度序列的统计特性呢？又如在测量陀螺漂移率时，测量系统也带有噪声，那么如何从测试记录数据中求出陀螺漂移率的模型呢？等等. 解决这些问题都要用到序列分解方法.

我们这里只讨论两个独立序列和的分解问题. 即使如此，如果不再加其它约束条件，一般来说，这种分解问题也得不到唯一的解. 例如 ARMA(1, 1) 序列，它既可以由一个 ARMA(1, 1) 序列与一白噪声合成；也可以由一个 AR(1, 0) 序列与一白噪声合成(详见后面的阐述). 由此可见，为使分解问题有意义，至少要假定每个分序列的模型阶数是已经给定的. 不难看出，即使在添加这项假定之下，也还不一定能给出唯一分解. 比如，一个白噪声序列 a_t，要把它分解成两个独立的白噪声之和，虽然这里每个分序列都给定了模型阶数（即都是 ARMA(0, 0) 型），然而所得到参数关系式只有 $\sigma_a^2=\sigma_1^2+\sigma_2^2$，因此仍得不到 σ_1^2 和 σ_2^2 的唯一解. 那么，在求解 ARIMA 序列的分解问题时，除了需要预先知道这两个分序列的模型阶数外，还要添加哪些条件，才能把每个分序列的模型参数定出来呢？这是本节要讨论的重点内容之一. 下面就不同情况分别加以研究.

首先分析自回归部分的可分解条件. 由合成模型与分序列模型之间的自回归部分的关系 (6.1.11) 式可知，

$$\varphi(B)=\varphi^{(1)}(B)\varphi^{(2)}(B),\ p=p_1+p_2$$

因此，为要从 $\varphi(B)$ 已知和 p_1，p_2 已知的条件下，唯一分解出 $\varphi^{(1)}(B)$ 和 $\varphi^{(2)}(B)$ 来，一般必须知道 $\varphi^{(i)}(B)$ $(i=1,2)$ 中的一个。或当 p_1 和 p_2 中有一个取零时，例如 $p_2=0$，$p_1=p$，这时只可能是 $\varphi^{(1)}(B)=\varphi(B)$，$\varphi^{(2)}(B)=1$.

其次讨论滑动平均部分的可分解条件. 设合成序列的滑动平均部分为 $\theta(B)a_t$，由上一节 (6.1.7) 式可知，如果能分别求出 $\Theta^{(1)}(B)$ 和 $\Theta^{(2)}(B)$，且 $\varphi^{(2)}(B)$ 及 $\varphi^{(1)}(B)$ 在上面的条件下是可分解的，那么，由于

$$\begin{aligned}\Theta^{(1)}(B)&=\varphi^{(2)}(B)\theta^{(1)}(B)\nabla^{d-d_1}\\ \Theta^{(2)}(B)&=\varphi^{(1)}(B)\theta^{(2)}(B)\nabla^{d-d_2}\end{aligned}\tag{6.2.1}$$

而 d，d_1，d_2 是已知，于是 $\theta^{(1)}(B)$，$\theta^{(2)}(B)$ 也就等于获得了. 因此关键在于如何确定出 (6.2.1) 式左方的两个 B 的多项式. 注意，由 (6.1.9) 式知 y_t 是 MA$(0,q)$ 序列，$y_t^{(i)}$ 是 MA$(0,q_i^*)$ 序列 $(i=1,2)$，其中

$$q_1^*=p_2+q_1+d-d_1,\quad q_2^*=p_1+q_2+d-d_2\tag{6.2.2}$$

再由 (6.1.5) 式便知道分序列 $y_t^{(i)}$ 的参数 (即 (6.2.1) 式的系数) 与 y_t 的参数的关系式，而由(6.1.7)式又可知 y_t 的模型为

$$y_t=\varphi^{(1)}(B)\varphi^{(2)}(B)\nabla^d z_t=\varphi(B)\nabla^d z_t=\theta(B)a_t\tag{6.2.3}$$

即合成序列的滑动平均部分 $\theta(B)$ 的诸系数恰为 y_t 的诸参数. 由 (6.2.1) 式可知，每个 $\Theta_j^{(i)}$ 都是 $\theta_1^{(i)},\theta_2^{(i)},\cdots,\theta_{q_i}^{(i)}$ $(i=1,2)$ 的线性函数，于是，代入 (6.1.5) 式便得到 $\boldsymbol{\theta}$ 与 $\boldsymbol{\theta}^{(1)}$，$\boldsymbol{\theta}^{(2)}$ 的关系式

$$\begin{cases}\sigma_a^2(1+\theta_1^2+\cdots+\theta_q^2)=\sigma_1^2f_0^{(1)}(\boldsymbol{\theta}^{(1)})+\sigma_2^2f_0^{(2)}(\boldsymbol{\theta}^{(2)})\\ \sigma_a^2(-\theta_k+\theta_1\theta_{k+1}+\cdots+\theta_{q-k}\theta_q)=\sigma_1^2f_k^{(1)}(\boldsymbol{\theta}^{(1)})+\sigma_2^2f_k^{(2)}(\boldsymbol{\theta}^{(2)})\end{cases}$$
$$k=1,2,\cdots,q\tag{6.2.4}$$

其中 $f_k^{(i)}(\boldsymbol{\theta}^{(i)})$ 都是 $\boldsymbol{\theta}^{(i)}$ 的二次多项式，$\sigma_i^2f_k^{(i)}(\boldsymbol{\theta}^{(i)})$ 恰是 $y_t^{(i)}$ 的自协方差函数. 至于能否由 (6.2.4) 式定出 $\boldsymbol{\theta}^{(1)}$ 和 $\boldsymbol{\theta}^{(2)}$ 及 σ_1^2 和 σ_2^2 来，可分以下三种情况进行讨论.

1. $q<q_1+q_2+1$ 时，即 (6.2.4) 式中方程个数小于未知量的个数，这时一般不能唯一定出分序列的未知参数.

2. $q = q_1 + q_2 + 1$ 时，即方程 (6.2.4) 式的未知量个数与方程个数一致．虽然这时一般能够唯一地解出 $\sigma_1^{(2)}$, $\theta^{(1)}$ 和 $\sigma_2^{(2)}$, $\theta^{(2)}$ 来，但是，为了使这些解属于相应的可逆域，即 $\theta^{(i)}(\lambda) = 0$ 的根都在单位圆外，而且 $\sigma_i^2 > 0$，还必须在 $\boldsymbol{\theta}$ 满足某些约束条件下才能实现有意义的分解．

3. $q > q_1 + q_2 + 1$ 时，即 (6.2.4) 式的未知量个数少于方程个数．依方程理论，(6.2.4) 式的 $q + 1$ 个方程中只能有 $q_1 + q_2 + 2$ 个是独立的，否则可能出现矛盾方程而没有解．因此只有对 θ 加上更苛刻的约束条件才可能分解．

综合上述可以看出，关于 ARIMA(p, d, q) 序列可分解条件的一般讨论是相当复杂和烦琐的．下面我们仅对几种低阶模型的分解条件加以具体说明．

例 1. 设 z_t 为 ARMA(1, 1) 序列，即

$$(1 - \varphi_1 B) z_t = (1 - \theta_1 B) a_t$$

其中 φ_1, θ_1, σ_a^2 已知，要把 z_t 分解为 AR(1, 0) 序列 $z_t^{(1)}$ 与白噪声序列 $a_t^{(2)}$ 之和，求其各参数 $\varphi_1^{(1)}$, σ_1^2 $(E(a_t^{(1)})^2 = \sigma_1^2)$ 和 $\sigma_2^2 (E(a_t^{(2)})^2 = \sigma_2^2)$ 之值．

首先由于 $q = 1$, $q_1 = q_2 = 0$, 所以 $q = q_1 + q_2 + 1$ 以及

$$\varphi_1^{(1)} = \varphi_1, \quad \Theta^{(1)}(B) = 1, \quad \Theta^{(2)}(B) = 1 - \varphi_1 B$$

将它们代入 (6.1.5) 式可得

$$\begin{cases} \sigma_a^2(1 + \theta_1^2) = \sigma_1^2 + \sigma_2^2(1 + \varphi_1^2) \\ \sigma_a^2 \theta_1 = \sigma_2^2 \varphi_1 \end{cases} \tag{6.2.5}$$

从此式容易求出 σ_1^2 和 σ_2^2 的解．但是为了保证它们都是正数，还必须要求 φ_1 和 θ_1 满足以下约束条件：

$$0 < \sigma_2^2 = \frac{\sigma_a^2 \theta_1}{\varphi_1}, \quad 0 < \sigma_1^2 = \sigma_a^2(1 + \theta_1^2) - \frac{\sigma_a^2 \theta_1 (1 + \varphi_1^2)}{\varphi_1}$$

即

$$\theta_1 \varphi_1 > 0, \qquad \frac{1 + \theta_1^2}{1 + \varphi_1^2} > \frac{\theta_1}{\varphi_1} \tag{6.2.6}$$

再加上 (φ_1, θ_1) 应当满足的平稳可逆性条件，即

$$|\varphi_1| < 1, \quad |\theta_1| < 1 \tag{6.2.7}$$

在 (θ_1, φ_1) 平面上绘出满足 (6.2.6) 和 (6.2.7) 两式的区域，图 6.2.1 中区域 A 和 B 即是. 为使 ARMA(1, 1) 序列能够唯一分解成 AR(1, 0) 序列与白噪声序列的独立和，参数 (θ_1, φ_1) 必须落在区域 A 或 B 中.

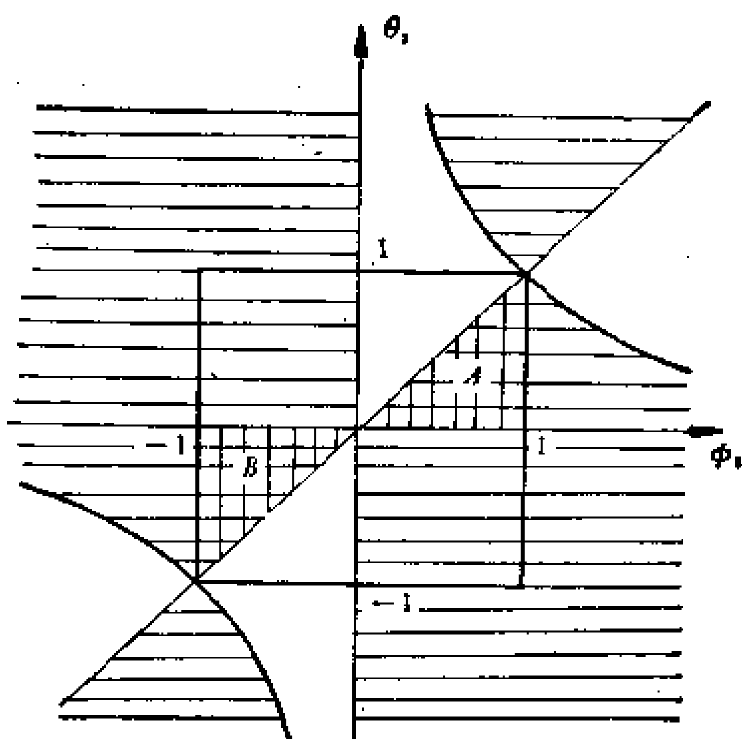


图 6.2.1 ARMA(1, 1) 序列可分解域

例2. 设 z_t 是 ARIMA(1, 1, 2) 序列，即

$$(1 - \varphi_1 B)\nabla z_t = (1 - \theta_1 B - \theta_2 B^2)a_t$$

其中 $\varphi_1, \theta_1, \theta_2$ 和 σ_a^2 是这个合成序列的诸参数. 要把它分解成 $z_t^{(1)}$ 与 $z_t^{(2)}$ 的独立和，其中 $z_t^{(1)}$ 为 AR(1, 0) 序列，$z_t^{(2)}$ 为 ARIMA(0, 1, 1) 序列，求 $z_t^{(i)}$ 的诸参数之值. 设

$$(1 - \varphi_1^{(1)} B)z_t^{(1)} = a_t^{(1)}$$

$$\nabla z_t^{(2)} = (1 - \theta_1^{(2)} B)a_t^{(2)}$$

容易知道，$\varphi_1^{(1)} = \varphi_1$，这时 (6.2.1) 式具体为

$$\Theta^{(1)}(B) = 1 - \Theta_1^{(1)}B = 1 - B$$

$$\Theta^{(2)}(B) = 1 - \Theta_1^{(2)}B - \Theta_2^{(2)}B^2 = (1 - \varphi_1 B)(1 - \theta_1^{(2)}B)$$

$$= 1 - (\varphi_1 + \theta_1^{(2)})B + \varphi_1\theta_1^{(2)}B^2$$

代入(6.2.4)式(或(6.1.5)式)可得

$$\begin{cases}\sigma_a^2(1+\theta_1^2+\theta_2^2) = 2\sigma_1^2 + \sigma_2^2\{1+(\varphi_1+\theta_1^{(2)})^2+(\varphi_1\theta_1^{(2)})^2\} \\ \sigma_a^2(-\theta_1+\theta_1\theta_2) = -\sigma_1^2-\sigma_2^2\{\varphi_1+\theta_1^{(2)}+\varphi_1\theta_1^{(2)}(\varphi_1+\theta_1^{(2)})\} \\ \qquad -\sigma_a^2\theta_2 = \sigma_2^2\varphi_1\theta_1^{(2)}\end{cases} \tag{6.2.8}$$

在本例中，$q=2$，$q_1=0$，$q_2=1$，$q=q_1+q_2+1$. 为了使得能有唯一合理的分解，合成序列 z_t 的诸参数应当使(6.2.8)式的解 σ_1^2, σ_2^2 为正数，而且 $|\theta_1^{(2)}|<1$；为此必须要求 φ_1，θ_1 和 θ_2 满足一定的约束条件，这里不去详细讨论了.

例 3. 设 z_t 为 ARMA(2, 2) 序列，即

$$(1-\varphi_1B-\varphi_2B^2)z_t=(1-\theta_1B-\theta_2B^2)a_t$$

其中诸参数为已知. 要将 z_t 分解为 AR(2, 0) 序列 $z_t^{(1)}$ 与白噪声序列 $z_t^{(2)}$ 的独立和，求分序列的参数 $\varphi_1^{(1)}$，$\varphi_2^{(1)}$，σ_1^2 和 σ_2^2. 首先由(6.2.1)式容易知道

$$\Theta^{(1)}(B)=1, \qquad \Theta^{(2)}(B)=1-\varphi_1B-\varphi_2B^2$$

代入(6.2.4)式(或(6.1.5)式)可得

$$\begin{cases}\sigma_a^2(1+\theta_1^2+\theta_2^2)=\sigma_1^2+\sigma_2^2(1+\varphi_1^2+\varphi_2^2) \\ \sigma_a^2(-\theta_1+\theta_1\theta_2)= \qquad \sigma_2^2(-\varphi_1+\varphi_1\varphi_2) \\ \sigma_a^2\theta_2=\sigma_2^2\varphi_2\end{cases} \tag{6.2.9}$$

又知 $q=2$，$q_1=0$，$q_2=0$，于是 $q>q_1+q_2+1$. 由(6.2.9)式可以看出，三个方程式只有两个未知数 σ_1^2 和 σ_2^2，为使它们有解，由后两方程分别解出的 σ_2^2 必须相等；又为使解有意义，必须 $\sigma_2^2>0$ 和 $\sigma_1^2>0$. 为此对参数 φ_1，φ_2 和 θ_1，θ_2 所加的约束条件应是

$$\begin{cases}\dfrac{\theta_2}{\varphi_2}=\dfrac{\theta_1(1-\theta_2)}{\varphi_1(1-\varphi_2)}>0 \\ \dfrac{1+\theta_1^2+\theta_2^2}{1+\varphi_1^2+\varphi_2^2}>\dfrac{\theta_2}{\varphi_2}\end{cases} \tag{6.2.10}$$

如果把 $(\theta_1, \theta_2, \varphi_1, \varphi_2)$ 的平稳可逆域当成四维欧氏空间的一个子区域，那么由于约束条件(6.2.10)式中包括有“等号”约束，参数 $(\theta_1, \theta_2, \varphi_1, \varphi_2)$ 一般只能在四维欧氏空间中的三维超曲面上取值. 我们用样本数据给出 z_t 的模型参数的估计值 $\hat{\beta}=(\hat{\varphi}_1, \hat{\varphi}_2, \hat{\theta}_1, \hat{\theta}_2)$

虽然依概率收敛到真值 β，但 $\ddot{\beta}$ 在 β 的邻域内可以随机取值，而且渐近多维正态分布，所以落在一个降维的超曲面上的概率是可忽略的．因此 (6.2.10) 式的条件是相当苛刻的．

为了克服这一困难，当合成序列的估计参数 $\hat{\beta}$ 不满足约束条件(如 (6.2.10) 式，(6.2.6) 式等)时，根据 $\hat{\beta}$ 在一定意义下收敛于 β，我们可以在 $\hat{\beta}$ 附近找一个与它靠近的、又满足约束条件的 $\hat{\beta}^*$ 来代替它．然后用第五章的方法对 $\hat{\beta}^*$ 进行检验，若能够接受，便用这组参数进行序列分解．若不能接受，问题就比较复杂了，因为这种否定原假设的原因太多了，其中有的因为原合成序列不是指定的两种分序列的独立和．因此，在检验不能接受 $\hat{\beta}^*$ 而 $\hat{\beta}^*$ 与 $\hat{\beta}$ 相差又远时，应当考虑其它阶数模型的分解的可能性．

由以上分析可以看到，序列的分解问题一般是难于解决的．而在实际中遇到较多的分解问题，往往是将 ARMA(p, q) 序列分解为 ARMA(p, q) 序列与白噪声序列 N_t 的独立和．即

$$z_t = z_t^{(1)} + N_t \tag{6.2.13}$$

前面提到过的陀螺漂移模型的测定问题，就属于这种分解，其中 $z_t^{(1)}$ 表示陀螺的真实漂移过程，N_t 表示测试设备的噪声．一般我们只能得到它们的合成序列 z_t 的样本值，而要估计的是 $z_t^{(1)}$ 的模型参数．我们可以先对 z_t 的模型进行识别，并作出参数估计，然后再根据本节介绍的方法进行分解．因为只有当 $q \geqslant q_1 + 1$ 时才可能进行唯一合理的分解，所以我们可以分别取 $z_t^{(1)}$ 为 ARMA$(p, 0)$，ARMA$(p, 1)$，$\cdots$，ARMA$(p, q-1)$ 等模型进行分解的尝试，从中选出合理的分解来．如果一个也选不到，则应拒绝 N_t 与 $z_t^{(1)}$ 独立或 N_t 为白噪声的假设，再考虑 N_t 为复杂一些的 ARMA 序列．

§3. 季节性模型

一、什么是季节性模型

在不少实际问题中，随机序列的变化包含很明显的周期性规

律．比如，我们在第一章§1中所举的例2就是一例．固定海域的海水温度的变化有一定的随机性，但是，逐年间又有很强的相似之处，即类似于年周期性的规律，这种规律是由于季节变化的原因所引起的．其它与自然现象有明确联系的时间序列，如气温、雨量、用电量、用水量等都有类似的规律．

本节里，通过对两个具体例子的分析，介绍关于含有季节性周期规律序列的特殊处理方法．

在绪论和第一章中，我们曾提到电力负荷的例子，如果以 z_t 表示 t 时刻的用电负荷量，以一小时为采样间隔，那么容易想象，序列 z_t 将包含24小时的周期性变化（见绪论中图三）．一般上午有一用电高峰，晚上又有一用电高峰；中午有一用电较少的低谷，深夜则出现一天用电最少的低谷．如果纵观各年用电量的数据序列，则又可以发现与四季相应的用电变化．而且，各星期间的用电波动情况也会有某种相似之处．这就是说，在长时间的用电量变化序列中，包含了多种周期性变化的趋势，此外还会有不断增长的趋势．对于这种较复杂的随机现象，有很多描述方法．例如，可以用三角函数与多项式函数的混合回归方法，把这些趋势性变化的部分用非随机函数来描述，然后再对其残余量用本书前面介绍的方法进行时序分析．又如，可以把一个周期长度等于 s 倍采样间隔的随机序列 z_t，看成是 s 个相互独立的 ARIMA 序列 z_{st}, z_{st+1}, $\cdots$, z_{st+s-1}($t=0, 1, 2, \cdots$) 交替出现的结果，其中每一序列的采样间隔都是原间隔的 s 倍．以用电量为例，考虑周期为24小时，我们可以把每天固定第 i 小时（如早上八点，即 $i=8$）的用电量 z_{24t+i}（t 表示第 t 天，$t=0, 1, 2, \cdots$）看做一个随机序列，并用通常的 ARIMA 模型描述它．这样，分别取 i 从第1小时到第24小时，就可以用24个 ARIMA 序列的交替出现来描述 z_t．但事实上，序列 z_{st}, z_{st+1}, $\cdots$, z_{st+s-1} 之间往往并不是独立的，所以用这种方法并不能完整地反映序列 z_t 的全部联系．为了全面研究它的统计性质，还可以把（z_{st}, z_{st+1}, $\cdots$, z_{st+s-1}）当作一个 s 维随机向量序列进行分析，但这样做又会在计算上带来很大的麻烦．

在本章中，我们将采用一类疏系数的 ARIMA 模型，即"季节性模型"．设 s 为一正整数，一个随机序列 z_t，如果它符合于以下的模型：

$$\Phi(B^s)\nabla_s^D z_t = \Theta(B^s)A_t \tag{6.3.1}$$

则称 z_t 是周期为 s 的季节性序列．(6.3.1) 式称为季节性模型，其中

$$\nabla_s = 1 - B^s, \qquad \nabla_s^D = (1 - B^s)^D$$

$$\Phi(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \cdots - \Phi_P B^{Ps}$$

$$\Theta(B^s) = 1 - \Theta_1 B^s - \Theta_2 B^{2s} - \cdots - \Theta_Q B^{Qs}$$

若 $D = 0$ 时，则还应考虑稍广一点的模型：

$$\Phi(B^s)z_t = C + \Theta(B^s)A_t \tag{6.3.2}$$

(6.3.1) 和 (6.3.2) 式中的 A_t 一般不必是白噪声，而可设它是另一 ARIMA(p, d, q) 序列：

$$\varphi(B)\nabla^d A_t = \theta(B)a_t \tag{6.3.3}$$

联合 (6.3.3) 和 (6.3.1) 式，便得到

$$\varphi(B)\Phi(B^s)w_t = \theta(B)\Theta(B^s)a_t, w_t = \nabla^d\nabla_s^D z_t \tag{6.3.4}$$

我们称它为乘积型的季节性模型，其阶数常用 $(p, d, q)\times(P, D, Q)_s$ 表示．我们可以将 (6.3.4) 式展开成通常的 ARIMA 形式．例如将

$$w_t = (1 - \theta_1 B)(1 - \Theta_1 B^s)a_t \tag{6.3.5}$$

展开成

$$w_t = (1 - \theta_1 B - \Theta_1 B^s + \theta_1\Theta_1 B^{s+1})a_t \tag{6.3.6}$$

它是一个 MA$(0, s + 1)$ 模型，但是其系数 $\theta_1, \theta_2, \cdots, \theta_{s+1}$ 中有很多等于零，具体地有 $\theta_1 = \theta_1$，$\theta_2 = \cdots = \theta_{s-1} = 0$，$\theta_s = \Theta_1$，$\theta_{s+1} = -\theta_1\Theta_1$． 可见，尽管 w_t 的阶数 $(s + 1)$ 可以很高，但除了 $\theta_1, \theta_s, \theta_{s+1}$ 之外其它系数取零值，而且其中 $-\theta_{s+1}$ 还是 θ_1 和 θ_q 之积，也就是说 (6.3.6) 式中实际上只有 θ_1 和 θ_s 是自由参数．有时为了方便还可以把 (6.3.6) 式改写成

$$w_t = (1 - \theta_1^* B - \theta_s^* B^s - \theta_{s+1}^* B^{s+1})a_t, \tag{6.3.7}$$

$$-\theta_{s+1}^* = \theta_1^*\theta_s^*$$

用同样方法，对于一般形式的 (6.3.4) 式，也可化成类似 (6.3.7) 式

这样的带约束条件的疏系数 ARIMA 模型:

$$\varphi^*(B)w_t = \theta^*(B)a_t \qquad (6.3.8)$$

其中约束条件由 (6.3.4) 式中的 $\varphi(\cdot)$, $\Phi(\cdot)$, $\theta(\cdot)$ 及 $\Theta(\cdot)$ 而定. w_t 称为疏系数 ARMA $(p+Ps, q+Qs)$ 序列.

二、季节性模型的时序分析方法

对于季节性模型的时序分析,除去阶数识别有某些特点外,参数估计、模型检验和改进等方法与一般 ARMA 序列基本相同, 现在,我们通过对以下实例的处理,阐述季节性模型的时序分析方法.

我们取得了某地区某年某月的用电负荷量 z_t 的数据. 采样以小时为间隔,总长度 $n=720$. 为了进行预报,我们要估计 z_t 的随机模型. 从经验上看,它可能含有24小时的周期. 一般的程序是,先分析 z_t 的样本自相关函数和样本偏相关函数; 若它们都不截尾

表 6.3.1 用电负荷序列的

序列	$\hat{\rho}_k$			(k 从 1 到 48)				$\gamma_0(z)=5720$,			
z_t	0.84	0.60	0.33	0.09	−0.13	−0.28	−0.35	−0.37	−0.33	−0.27	−0.13
	0.86	0.75	0.53	0.28	0.05	−0.16	−0.30	−0.37	−0.38	−0.33	−0.23
∇z_t	0.27	0.08	−0.08	−0.07	−0.22	−0.23	−0.20	−0.15	−0.21	0.03	0.17
	0.69	0.36	0.09	−0.06	−0.05	−0.23	−0.23	−0.18	−0.16	−0.18	0.00
$\nabla_{24} z_t$	0.58	0.44	0.35	0.27	0.27	0.21	0.14	0.13	0.05	0.05	0.04
	−0.52	−0.28	−0.24	−0.19	−0.16	−0.18	−0.13	−0.11	−0.08	−0.05	−0.05
$\nabla\nabla_{24} z_t$	−0.30	−0.06	−0.02	−0.09	0.07	0.01	−0.06	0.08	−0.09	0.01	−0.02
	−0.48	0.23	−0.01	−0.02	0.07	−0.09	0.03	−0.01	−0.00	0.04	−0.06

也不拖尾，再分析数据列 $\nabla_{24}z_t = z_t - z_{t-24}$ 的样本自相关与偏相关函数；若仍不截尾也不拖尾，可由低到高再考虑数据列 $\nabla\nabla_{24}z_t$，$\nabla\nabla_{24}^2 z_t$ 以及 $\nabla^2\nabla_{24}z_t$ 等. 一旦从中找到一个，可以认为它的样本自相关与偏相关具有截尾或拖尾性质后，就以这样的 $\nabla^d\nabla_{24}^D z_t = w_t$ 当做某个 ARMA 序列，以后的分析和估计便采用第三、四、五章的方法进行了. 按照这一手续，我们用 720 个 z_t 的样本值先后计算了 z_t，∇z_t，$\nabla_{24}z_t$，$\nabla\nabla_{24}z_t$ 的样本自相关函数 $\hat{\rho}_k$ 与偏相关函数 $\hat{\varphi}_{kk}$，并把它们的 $\hat{\rho}_k$ 值列在表 6.3.1 里. 由这些数据可以看出，z_t，∇z_t 的样本自相关函数既不象拖尾，也不象截尾，它们的值不趋于零，或者只是很慢地收敛于零. 而 $\nabla\nabla_{24}z_t$ 的 $\hat{\rho}_k$ 和 $\hat{\varphi}_{kk}$ 都近于拖尾的，其 $\hat{\rho}_k$ 值中只是 $\hat{\rho}_1$，$\hat{\rho}_{24}$ 较大，$\hat{\rho}_{23}$，$\hat{\rho}_{25}$ 其次，其余的 $\hat{\rho}_k$ 都近于零值. 这种样本自相关函数与(6.3.6)式的序列 w_t 的自相关函数非常相似，在(6.3.6)式中若以 24 代 s，则其自相关函数为

各种样本自相关函数

$r_0(\nabla z) = 1812$，　　$r_0(\nabla_{24}z) = 1280$，　　$r_0(\nabla\nabla_{24}z) = 1063$

−0.09	−0.14	−0.24	−0.34	−0.39	−0.37	−0.30	−0.16	0.06	0.28	0.53	0.75
−0.14	−0.10	−0.15	−0.24	−0.34	−0.38	−0.36	−0.28	−0.14	0.06	0.28 0.74	0.53 0.84
0.28	0.16	0.03	−0.19	−0.19	−0.17	−0.23	−0.23	−0.05	−0.07	0.10	0.35
0.18	0.28	0.14	0.03	−0.19	−0.17	−0.21	−0.20	−0.19	−0.07	−0.08 0.34	0.10 0.68
0.04	0.06	0.06	0.04	−0.02	−0.01	−0.09	−0.15	−0.13	−0.20	−0.25	−0.33
−0.01	−0.00	−0.04	−0.01	−0.02	−0.02	−0.03	0.04	0.06	0.00	0.02 0.06	0.04 0.05
−0.01	0.02	0.02	0.05	−0.09	0.10	−0.02	−0.09	0.10	−0.02	0.03	0.14
0.05	0.05	−0.08	0.05	−0.02	0.01	−0.08	0.05	0.10	−0.09	0.00 0.04	0.00 0.05

$$
\begin{cases}
\rho_1 = \dfrac{-\theta_1(1+\Theta_1^2)}{1+\theta_1^2+\Theta_1^2+\theta_1\Theta_1^2} = \dfrac{-\theta_1}{1+\theta_1^2} \\
\rho_{23} = \rho_{25} = \dfrac{\theta_1\Theta_1}{(1+\theta_1^2)(1+\Theta_1^2)} = \rho_1\rho_{24} \\
\rho_{24} = \dfrac{-\Theta_1-\theta_1^2\Theta_1}{(1+\theta_1^2)(1+\Theta_1^2)} = \dfrac{-\Theta_1}{1+\Theta_1^2} \\
\rho_k = 0,\ k \neq 0,\ 1,\ 23,\ 24,\ 25
\end{cases}
\tag{6.3.9}
$$

如要仔细检验 $\hat{\rho}_k(k>25)$ 的真值 ρ_k 是否可认为零，可用 Bartlett 公式检验，

$$
\mathrm{var}(\hat{\rho}_k) \approx \frac{1}{n}[1+2(\hat{\rho}_1^2+\hat{\rho}_{23}^2+\hat{\rho}_{24}^2+\hat{\rho}_{25}^2)] \tag{6.3.10}
$$

以表 6.3.1 中 $\nabla\nabla_{24}z_t$ 的 $\hat{\rho}_k$ 值代入上式得 $\mathrm{var}(\hat{\rho}_k)\cong 0.00243$ 则 $[\mathrm{var}(\hat{\rho}_k)]^{\frac{1}{2}}\cong 0.05$；又由于 $\hat{\rho}_k$ 渐近正态分布，将 k 从26到48的 $|\hat{\rho}_k|$ 值落在 $[0,\ 0.05]$ 和 $[0,\ 0.10]$ 中的比例求出，它们分别是

$$
\frac{15}{23} = 65.2\%, \qquad \frac{23}{23} = 100\%
$$

这与正态分布 $\mathscr{N}(0,(0.05)^2)$ 分别落在 $\pm(0.05)$，$\pm 2(0.05)$ 之间的概率 68.3%，95.5% 是比较接近的．由此可以粗略地判断 $\hat{\rho}_k$ 是零值的估计量．因此，我们可以用乘积型的 $(0,\ 1,\ 1)\times(0,\ 1,\ 1)_{24}$ 季节性模型来拟合用电量序列 z_t，即

$$
\nabla\nabla_{24}z_t = (1-\theta_1 B)(1-\Theta_1 B^{24})a_t \tag{6.3.11}
$$

为了求出其参数的矩估计，只须将 $\nabla\nabla_{24}z_t$ 的 $\hat{\rho}_k$ 代替 (6.3.9) 式的 ρ_k 便可解出 $\hat{\theta}_1$ 和 $\hat{\Theta}_1$，即由方程

$$
\hat{\rho}_1 = -0.3 = \frac{-\theta_1}{1+\theta_1^2}, \qquad \hat{\rho}_{24} = -0.48 = \frac{-\Theta_1}{1+\Theta_1^2}
$$

解得 $\hat{\theta}_1\cong 0.33$，$\hat{\Theta}_1\cong 0.75$．σ_a^2 的矩估计则为

$$
\hat{\sigma}_a^2 = \hat{\gamma}_0(1+\hat{\theta}_1^2)^{-1}(1+\hat{\Theta}_1^2)^{-1} \approx 1062\times(1+0.11)^{-1}
$$
$$
(1+0.56)^{-1} \approx 611 \cong (24.6)^2
$$

在矩估计的基础上，还可以进行精估计．因 $n=720$，对序列 $w_t=\nabla\nabla_{24}z_t$ 而言，样本长度 $N=720-25=695$．这个长度算是较大的，可以采用最小二乘估计作为精估计，即求使

$$S_N(\theta_1, \Theta_1) = \sum_{t=1}^{N} \hat{a}_t^2$$

达极小值的估计 $\hat{\theta}_1$ 和 $\hat{\Theta}_1$，其中 $\hat{a}_t = \hat{a}_t(\theta_1, \Theta_1)$ 可由如下的递推式计算：

$$\hat{a}_t = w_t + \theta_1 \hat{a}_{t-1} + \Theta_1 \hat{a}_{t-24} - \theta_1 \Theta_1 \hat{a}_{t-25}$$

而初始值 $\hat{a}_t = 0 (t \leqslant 0)$. 用图解法求出 $S_N(\theta_1, \Theta_1)$ 取极小值的解为 $\hat{\theta}_1 = 0.55$，$\hat{\Theta}_1 = 0.88$（格点上数值见表 6.3.2），相应的 $\hat{\sigma}_a^2 = (22.7)^2$. 当然也可以用迭代法求其极值解，结果几乎完全相同.

表 6.3.2(a) 粗 格 点

Θ_1 \ S_N \ θ_1	0.40	0.45	0.50	0.55	0.60	0.65
0.95	374706	371204	368921	367931	368348	370350
0.90	359880	356757	354811	354114	354782	356994
0.85	360297	357226	355326	354671	355380	357637
0.80	365600	362494	360568	359897	360605	362882
0.75	373136	369965	367991	367291	367993	370298

表 6.3.2(b) 细 格 点

Θ_1 \ S_N \ θ_1	0.53	0.54	0.55	0.56	0.57	0.58
0.90	354237	354149	354114	354134	354209	354341
0.89	353780	353624	353593	353616	353694	353830
0.88	353567	353485	353456	353481	353562	353700
0.87	353725	353654	353627	353653	353736	353875
0.86	354153	354073	354046	354074	354157	354297

对于模型（6.3.11）式的检验，可以用第五章所介绍的各种方法. 例如先求出模型

$$\nabla \nabla_{24} z_t = (1 - 0.55B)(1 - 0.88B^{24}) a_t \qquad (6.3.12)$$

的"残量" a_t 的样本值 $\hat{a}_t$，及其样本自相关函数 $\hat{\rho}_k(\hat{a})$，然后计算

$$Q_{48} = N \sum_{k=1}^{48} \hat{\rho}_k^2(\hat{a}) = 695 \sum_{k=1}^{48} \hat{\rho}_k^2(\hat{a}) = 62.6$$

Q_{48} 渐近 χ^2 分布，其自由度为 48. 我们取显著水平 $\alpha = 0.05$，由 χ^2 表查出

$$P(Q_{48} > 65.2) = 0.05$$

这样，由于计算的样本值 $Q_{48} = 62.6 < 65.2$，因此可以接受(6.3.12)式的模型．鉴于一般统计书中的 χ^2 分布表以低自由度的为主，而时间序列分析中的检验问题多用高自由度的 χ^2 分布，我们特在本书末尾列出高自由度的 χ^2 表供读者使用．

如果我们把表 6.3.1 中 $\nabla_{24}z_t$ 的 $\hat{\rho}_k$ 看成是拖尾的，那么还可以用另一种模型来拟合上述的用电序列．令 $w_t = \nabla_{24}z_t = z_t - z_{t-24}$，试以它适合模型

$$(1-\varphi_1 B)(1-\Phi_1 B^{24})w_t = a_t \qquad (\Phi_1 < 0) \qquad (6.3.13)$$

w_t 的理论自相关函数满足

$$(1-\varphi_1 B)(1-\Phi_1 B^{24})\rho_k = 0, \qquad k > 0 \qquad (6.3.14)$$

由此式解出 ρ_k 的一般形式为

$$\rho_k = c_0\varphi_1^k + |\Phi_1|^{\frac{k}{24}} \sum_{j=1}^{12} \left(c_j' \cos \frac{k(2j+1)\pi}{24} + c_j'' \sin \frac{k(2j+1)\pi}{24} \right), \qquad k > -25 \qquad (6.3.15)$$

与表 6.3.1 中 $\nabla_{24}z_t$ 的 $\hat{\rho}_k$ 对照，当 k 从 0 变到 36 时，形状是比较相似的，当 $k > 36$ 以后就拟合得不太好了．所以，用(6.3.13)式的模型来拟合所给的用电序列，可能比用(6.3.11)式要差些．但是，如果建立模型的目的只是为了进行预报，那么用(6.3.13)式的模型比(6.3.11)式要来得方便(见第七章)．因此，如果两者的预报效果相差不多．还是用(6.3.13)式好．下面就来分析这种模型的参数估计的具体情况．

利用(6.3.14)式，将 ρ_k 代之以 $\nabla_{24}z_t$ 的 $\hat{\rho}_k$，可以获得 φ_1，Φ_1 的一种矩估计方程：

$$\begin{cases} \hat{\rho}_1 = \hat{\varphi}_1 + \hat{\Phi}_1\hat{\rho}_{23} - \widehat{\varphi_1\Phi_1}\hat{\rho}_{24} \\ \hat{\rho}_{24} = \hat{\varphi}_1\hat{\rho}_{23} + \hat{\Phi}_1 - \widehat{\varphi_1\Phi_1}\hat{\rho}_1 \\ \hat{\rho}_{25} = \hat{\varphi}_1\hat{\rho}_{24} + \hat{\Phi}_1\hat{\rho}_1 - \widehat{\varphi_1\Phi_1} \end{cases} \qquad (6.3.16)$$

这里我们把式中的 $\widehat{\varphi_1\Phi_1}$ 做为一个独立的未知元，因此 (6.3.16) 式为三元一次联立方程. 由本节表 6.3.1 查出

$$\hat{\rho}_1 = 0.58, \quad \hat{\rho}_{23} = -0.33, \quad \hat{\rho}_{24} = -0.52$$
$$\hat{\rho}_{25} = -0.28, \quad \hat{\gamma}_0(\nabla_{24}z_t) = 1280$$

并注意 $\sigma_a^2 = Ea_t^2 = E(w_t - \varphi_1 w_{t-1} - \Phi_1 w_{t-24} + \varphi_1\Phi_1 w_{t-25})^2$，式中的 $\varphi_1, \Phi_1, \gamma_0, \rho_k$ 都以估计值代替，于是解出：

$$\hat{\varphi}_1 = 0.56, \ \hat{\Phi}_1 = -0.52, \ \widehat{\varphi_1\Phi_1} = -0.31, \ \hat{\sigma}_a^2 = (25.2)^2$$

由此可见，以 $\widehat{\varphi_1\Phi_1}$ 做为 $\varphi_1\Phi_1$ 的估计，和以 $\hat{\varphi}_1 \cdot \hat{\Phi}_1 = 0.56(-0.52) \cong -0.29$ 做为 $\varphi_1\Phi_1$ 的估计是比较接近的. 因此，我们可以接受 $\hat{\varphi}_1 = 0.56$, $\hat{\Phi}_1 = -0.52$ 的估值. 如果 $\hat{\varphi}_1\hat{\Phi}_1$ 与 $\widehat{\varphi_1\Phi_1}$ 相差较远，可能因为 (6.3.13) 式的模型不合适的缘故.

如果在应用中发生这种情况，我们不妨放弃约束 $\Phi_2 = \varphi_1\Phi_1$，令其为自由参数，而采用以下的拟合模型:

$$w_t - \varphi_1 w_{t-1} - \Phi_1 w_{t-24} - \Phi_2 w_{t-25} = a_t \qquad (6.3.17)$$

这里我们不去分析这类拟合问题了.

在求 (6.3.13) 式的精估计时，由于带有约束条件，即使采用最小二乘估计，也要遇到二元高次联立方程，所以只能用数值解法. 我们定义

$$S(\varphi_1, \Phi_1) = \sum_{t=26}^{720} \hat{a}_t^2$$

其中 a_t 由如下递推式计算:

$$\hat{a}_t = w_t - \varphi_1 w_{t-1} - \Phi_1 w_{t-24} + \varphi_1\Phi_1 w_{t-25}$$

初始值 $\hat{a}_t = 0\ (t \leqslant 0)$，用图解法求出 $S(\varphi_1, \Phi_1)$ 的极小值解，其结果为 (见表 6.3.3)

$$\hat{\varphi}_1 = 0.56, \quad \hat{\Phi}_1 = -0.52, \quad \hat{\sigma}_a^2 = (25.2)^2$$

与矩估计完全一样.

如上所述，我们求得 z_t 的另一种可取的模型估计为

$$(1 - 0.56B)(1 + 0.52B^{24})\nabla_{24}z_t = a_t \qquad (6.3.18)$$

由 $\hat{\sigma}_a^2$ 的大小可以看出，用这一模型比用 (6.3.12) 式模型的拟合

表 6.3.3(a) 粗格点

Φ \ S_N \ φ	0.52	0.56	0.60	0.64	0.68
−0.44	446530	445497	446541	449611	454858
−0.48	443868	442890	443974	447118	452324
−0.52	443070	442141	443266	446446	451680
−0.56	444137	443248	444417	447644	452927

表 6.3.3(b) 细格点

Φ \ S_N \ φ	0.54	0.55	0.56	0.57	0.58
−0.50	442503	442329	442283	442360	442578
−0.51	442367	442196	442154	442240	442454
−0.52	442349	442180	442141	442229	442446
−0.53	442446	442280	442243	442335	442554
−0.54	442660	442497	442462	442556	442779

“残差”要大，甚至比(6.3.11)式的矩估计模型的拟合“残差”还大．根据第五章的方法，计算(6.3.18)式拟合“残差”a_t的样本值$\hat{a}_t$，并求出了它的样本自相关函数$\hat{\rho}_k(\hat{a})$．如以前一样，计算出统计量

$$Q_{48}=N\sum_{k=1}^{48}\hat{\rho}_k^2(\hat{a})=695\sum_{k=1}^{48}\hat{\rho}_k^2(\hat{a})=148.9$$

仍以0.05为显著性水平，经查表得到$P(Q_{48}>65.2)=0.05$，而148.9远远大于65.2．因此，一般说来不能接受(6.3.18)式的模型．

我们再举一个例子，这是在研究地球自转稳定性时所遇到的数据处理问题．在第一章§1的例3中，已经提到过地球自转一周时间的起伏，我们希望用ARIMA模型来拟合这种随机变化量，以供预报或其它方面应用．

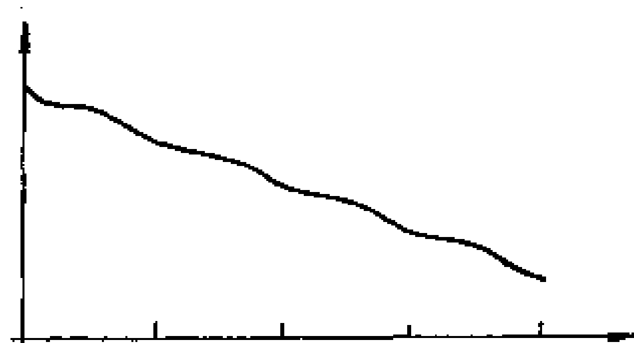


图 6.3.1

图 6.3.1 给出了1963年 3 月 26 日至1 967 年 3 月 25 日（共四年）地球自转随时间变化的情况，共 $N = 293$ 个数据（每五天取一个）．从图6.3.1中可以看出明显的趋势性变化．对数据更仔细的观察还看到有一年为周期的变化．我们算出了 $w_t = \nabla\nabla_{73}z_t$（这里 73 恰好是一年的取样个数）的样本自相关函数 $\hat{\rho}_k$ 和偏相关函数 $\hat{\varphi}_{kk}$，列在表 6.3.4 中，其中 $\hat{\gamma}_0 = 226.0$．按照第三章 §2 的模型初步识别方法，从表四的诸数值可以看出，$\hat{\rho}_k$ 是逐渐下降的，$\hat{\varphi}_{kk}$ 近于截尾，可以初步确定 w_t 的模型为 AR(1, 0) 或 AR(2, 0)．

表 6.3.4

	k 从 1 至33				($\hat{\gamma}_0 = 226$)						
$\hat{\rho}_k$	0.956	0.891	0.813	0.735	0.683	0.644	0.618	0.604	0.591	0.586	0.583
	0.585	0.586	0.590	0.594	0.588	0.584	0.580	0.584	0.596	0.607	0.623
	0.631	0.633	0.622	0.596	0.565	0.528	0.494	0.499	0.468	0.470	0.479
$\hat{\varphi}_{kk}$	0.956	−0.265	−0.141	0.007	0.317	−0.044	0.016	0.080	0.034	0.082	0.019
	0.172	−0.033	0.112	0.017	−0.072	0.074	0.108	0.135	0.003	0.003	…

若采用 AR(1, 0) 模型，其参数估计为 $\hat{\varphi}_1 = 0.956$，$\sqrt{\hat{\sigma}_a^2} = \sqrt{\hat{\gamma}_0(1 - \hat{\varphi}_1^2)} = 4.4$毫秒，于是 z_t 的估计模型为

$$(1 - 0.956B)\nabla\nabla_{73}z_t = a_t \tag{6.3.19}$$

由此可见，一般当 $\nabla^d\nabla_T^D z_t = w_t$ 能够用低阶 ARMA 模型拟合时，这种季节性模型并不表现出疏系数的形式(即 $(p, d, q)\times(P, D, Q)_T$ 中的$P = Q = 0$)，只是对 z_t 作差分运算时，含有以周期为步长的差分形式而已．若能用这种模型拟合，w_t 的参数估计与非季节性模型情况完全相同，其检验手续也相同，这里不一一列举了．

第七章　随机序列预报

§1. 预备公式

我们在这一章里着手研究随机序列预报的数学方法．前面多次提到的电力负荷的例子，海水温度的例子，以及地球自转速率变化的例子等，都要求进行预报．在气象、水文、地理和天文等方面有很多类似这样的问题，它们要求根据以往积累的数据进行时序分析，确定出其随机模型，然后对未来可能出现的结果进行预报．

本节的目的，是为 ARIMA 序列预报的数学方法做准备，推导一些对今后有用的公式．本节限于讨论均值为零的正态 ARIMA 序列，而且统一用 w_t 代表 ARMA 序列，z_t 代表 ARIMA 序列．所谓 l 步预报，就是根据现在时刻 k 和以前各时刻的序列值 w_k，w_{k-1}，$\cdots$，对未来时刻 $k+l$ 的随机变量 w_{k+l} 作出估计，因此，我们需要第二章的各种模型定义和公式，从中得出 w_{k+l} 的各种表达式．

直接形式：在 (2.1.5) 式中将 t 代以 $k+l$，并移动有关项可得

$$w_{k+l}=\varphi_1 w_{k+l-1}+\varphi_2 w_{k+l-2}+\cdots+\varphi_p w_{k+l-p}-\theta_1 a_{k+l-1}\\-\theta_2 a_{k+l-2}-\cdots-\theta_q a_{k+l-q}+a_{k+l}\tag{7.1.1}$$

逆转形式：在 (2.1.23) 式中将 t 代以 $k+l$，并移动有关项可得

$$w_{k+l}=a_{k+l}+\sum_{j=1}^{\infty}\pi_j w_{k+l-j}\tag{7.1.2}$$

传递形式：在 (2.1.17) 式中将 t 代以 $k+l$ 便得

$$w_{k+l}=\sum_{j=0}^{\infty}\phi_j a_{k+l-j}=\sum_{j=k+l}^{\infty}\phi_{k+l-j}a_j\tag{7.1.3}$$

分解形式：在 (7.1.3) 式中，将等式右边的求和项分成两部

分，可得

$$w_{k+l}=a_{k+l}+\psi_1 a_{k+l-1}+\cdots+\psi_{l-1}a_{k+1}+\sum_{j=l}^{\infty}\psi_j a_{k+l-j}$$
$$=D_k(l)+C_k(l) \tag{7.1.4}$$

其中

$$D_k(l)=\begin{cases}\sum_{j=0}^{l-1}\psi_j a_{k+l-j}, & l>0\\ 0\ , & l\leqslant 0\end{cases} \tag{7.1.5}$$

$$C_k(l)=w_{k+l}-D_k(l)=\begin{cases}\sum_{j=l}^{\infty}\psi_j a_{k+l-j}, & l>0\\ w_{k+l}\ , & l\leqslant 0\end{cases} \tag{7.1.6}$$

以上四种表达形式在推导预报公式中非常有用．其中传递形式和分解形式表面上看来完全相同，好象只是在分解形式里把一个无穷求和分成两个部分． 但是，$C_k(l)$ 和 $D_k(l)$ 具有很重要的性质和作用，其中 $C_k(l)$ 称为预报函数． 下面讨论 $D_k(l)$ 和 $C_k(l)$ 的性质．

为了方便起见，我们记 $k+l=t$，由 $l=t-k$，于是有

$$w_t=C_k(t-k)+D_k(t-k) \tag{7.1.7}$$

将 k 固定，以延迟算子 B 作用于 t（也就是作用于 l）有

$$\theta(B)a_t=\varphi(B)w_t=\varphi(B)C_k(t-k)+\varphi(B)D_k(t-k)$$

当 $0\leqslant i\leqslant p$ 时，

$$B^iC_k(t-k)=C_k(t-i-k)=C_k(l-i)$$
$$=\sum_{j=l-i}^{\infty}\psi_j a_{k+l-i-j}=\sum_{s=0}^{\infty}\psi_{l-i+s}a_{k-s}$$

其中约定 $\psi_j=0$，$j<0$，于是

$$\varphi(B)C_k(t-k)=\sum_{s=0}^{\infty}(\psi_{l+s}-\varphi_1\psi_{l-1+s}-\cdots-\varphi_p\psi_{l-p+s})a_{k-s}$$
$$=\sum_{s=0}^{\infty}[\varphi(B)\psi_{l+s}]a_{k-s}$$

根据（2.1.21）式之前面的公式，当 $l>q$ 时，$\varphi(B)\psi_{l+s}=0$，所以

$$\varphi(B)C_k(t-k)=0, \qquad t-k>q \qquad (7.1.8)$$

从而有

$$\varphi(B)D_k(t-k)=\theta(B)a_t, \qquad t-k>q \qquad (7.1.9)$$

这表明 $D_k(t-k)$ 是差分方程

$$\varphi(B)w_t=\theta(B)a_t, \qquad t-k>q \qquad (7.1.10)$$

的一个解，而 $C_k(t-k)$ 则是差分方程 (7.1.8) 式的一个解. 但 (7.1.8) 式的通解可表为

$$C_k(t-k)=\sum_{i=1}^{M}\left[\sum_{r=0}^{r_i-1} b_{ir}^{(k)}(t-k)^r\right]\omega_i^{-(t-k)}, \quad t-k>q-p \qquad (7.1.11)$$

其中 ω_i 是 $\varphi(\omega)=0$ 的 r_i 重根，对于不同的 i，ω_i 不相同，且 $r_1+r_2+\cdots+r_M=p$. 若 $\varphi(\omega)=0$ 无重根，则通解的形式可简化成

$$C_k(t-k)=\sum_{i=1}^{p} d_i^{(k)}\omega_i^{-(t-k)}, \quad t-k>q-p \quad (7.1.12)$$

ω_i 可能为实根或共轭复根，所以又可以把 (7.1.11) 式 (或 (7.1.12) 式) 写成如下形式:

$$C_k(t-k)=\sum_{k=1}^{p} b_i^{(k)}f_i(t-k), \quad t-k>q-p \quad (7.1.13)$$

其中 $f_i(t-k)$ 由 $(t-k)$ 的指数函数、三角函数和幂函数的乘积构成，而 $b_i^{(k)}$ 由初始值 $C_k(q)$, $C_k(q-1)$, $\cdots$, $C_k(q-p+1)$ 确定. 这些初始值可以按 (7.1.6) 式计算. 可见，$b_i^{(k)}$ 是由 $a_j(j\leqslant k)$ 的加权线性和组成的，它们都是随机量.

例如 ARMA(1, 1) 序列的 ϕ_j 已由 (2.1.19) 式给出，从此代入 (7.1.6) 式可知，当 $l=t-k>0$ 时

$$\begin{aligned} C_k(t-k) &= \sum_{j=l}^{\infty}\phi_j a_{t-j}=\sum_{j=l}^{\infty}\varphi^{j-1}(\varphi-\theta)a_{t-j} \\ &= \left[\sum_{s=0}^{\infty}\varphi^s(\varphi-\theta)a_{k-s}\right]\varphi^{l-1}=b^{(k)}\varphi^{t-k} \end{aligned}$$

其中

$$b^{(k)} = \frac{\varphi - \theta}{\varphi} \sum_{s=0}^{\infty} \varphi^s a_{k-s} = \frac{1}{\varphi} (\varphi w_k - \theta a_k)$$

§2. 平稳线性最小方差预报及其性质

在这一节与下一节，我们都用 $\hat{w}_k(l)$ 表示根据序列的历史值 $w_k, w_{k-1}, \cdots$，对未来 w_{k+l} 的预报值. 当然，在实际问题中并不可能知道全部历史值，而只能知道有限个历史值；我们这样定义预报值，是为了给出比较简单的预报公式和方法. 然而，当历史数据 $w_k, w_{k-1}, \cdots, w_1$ 的个数足够多时，即 k 很大以后，我们将看到，用全部历史值预报与用 k 个历史值预报的效果几乎完全一样.

那么，$\hat{w}_k(l)$ 是怎样定义的呢？也就是说，依据什么原则进行预报呢？我们采用前几章用过的线性最小方差估计的方法来定义 $\hat{w}_k(l)$. 根据第一章 §5 的内容，引进以下记号:

$$\mathscr{W}_k = \left\{w: \ w = \sum_{j=-\infty}^{k} c_j w_j, \ \sum_{j=-\infty}^{k} c_j^2 < \infty\right\}$$

$$\mathscr{A}_k = \left\{a: \ a = \sum_{j=-\infty}^{k} d_j a_j, \ \sum_{j=-\infty}^{k} d_j^2 < \infty\right\}$$

$$\hat{w}_k(l) = E(w_{k+l} | w_k, w_{k-1}, \cdots) = E(w_{k+l} | \mathscr{W}_k)$$

应用第一章 §5 中关于最小方差估计的各条性质，再根据上一节关于 w_{k+l} 的四种表达形式，可知

1. $\hat{w}_k(l) = E(w_{k+l} | \mathscr{W}_k) = E(w_{k+l} | \mathscr{A}_k)$; $\mathscr{W}_k = \mathscr{A}_k$;
2. $\hat{w}_k(-l) = w_{k-l}, \ l \geqslant 0$;
3. $\hat{a}_k(l) = E(a_{k+l} | \mathscr{W}_k) = E(a_{k+l} | \mathscr{A}_k)$;
4. $\hat{a}_k(-l) = a_{k-l}, \ l \geqslant 0$; $\hat{a}_k(l) = 0, \ l > 0$;
5. $\hat{w}_k(l) = \varphi_1 \hat{w}_k(l-1) + \varphi_2 \hat{w}_k(l-2) + \cdots$

$$+ \phi_p \hat{w}_k(l-p), \ l > q \qquad (7.2.1)$$

此式称为预报的差分方程式. 如果 w_t 为 MA 序列，则 (7.2.1) 式变为

$$\hat{w}_k(l) = 0, \qquad l > q \qquad (7.2.2)$$

6. $\hat{w}_k(l)=\sum_{j=1}^{l-1}\pi_j\hat{w}_k(l-j)+\sum_{j=l}^{\infty}\pi_j w_{k+l-j},\quad l>0$ (7.2.3)

因为 $\hat{w}_k(l-j)$ $(j=1,2,\cdots,l-1)$ 也是 $w_{k-j}(j=0,1,\cdots)$ 的线性和,所以可令

$$\hat{w}_k(l)=\sum_{j=1}^{\infty}\pi_j^{(l)}w_{k+1-j}=\pi^{(l)}(B)w_k \tag{7.2.4}$$

其中 $\pi^{(l)}(B)=\sum_{j=1}^{\infty}\pi_j^{(l)}B^{j-1}$. 将(7.2.4)式代入(7.2.3)式,并对比两边系数,容易证明

$$\begin{cases}\pi_j^{(l)}=\pi_{j+l-1}+\sum_{i=1}^{l-1}\pi_i\pi_j^{(l-i)}, & l>1\\ \pi_j^{(1)}=\pi_j\end{cases} \tag{7.2.5}$$

7. $\hat{w}_k(l)=C_k(l)$ (7.2.6)

当 l 固定而 k 变化时,由(7.1.6)容易看出 $C_k(l), k=1,2,\cdots$ 是平稳序列. 因此,我们称这种预报为平稳线性最小方差预报.

再利用(7.1.13)式(或(7.2.1)式)还可得

$$\hat{w}_k(l)=C_k(l)=\sum_{i=1}^{p}b_i^{(k)}f_i(l),\qquad l>q-p \tag{7.2.7}$$

$b_i^{(k)}$ 由初值 $\hat{w}_k(q),\hat{w}_k(q-1),\cdots,\hat{w}_k(q-p+1)$决定,即 $b_i^{(k)}$ 满足以下线性方程组

$$\sum_{i=1}^{p}b_i^{(k)}f_i(j)=\hat{w}_k(j),\quad j=q,q-1,\cdots,q-p+1 \tag{7.2.8}$$

8. $e_k(l)=w_{k+l}-\hat{w}_k(l)=D_k(l)$ (7.2.9)

称为 l 步预报误差. 特别当 $l=1$ 时,

$$e_k(1)=w_{k+1}-\hat{w}_k(1)=a_{k+1} \tag{7.2.10}$$

9. $\mathrm{var}[e_k(l)]=Ee_k^2(l)=(1+\psi_1^2+\cdots+\psi_{l-1}^2)\sigma_a^2$

$l>0$ (7.2.11)

这就是 l 步预报误差的方差. 由此看出, $E[\hat{w}_k(l)-w_{k+l}]^2=\mathrm{var}[e_k(l)]$ 只与 l 有关,而与 k 无关,这是平稳序列的平稳预报所特有的性质之一.

利用 (7.1.6), (7.2.6) 和 (7.2.10) 式还可得

10. $\hat{w}_{k+1}(l) = C_{k+1}(l) = \sum_{j=0}^{\infty} \psi_{l+j}a_{k+1-j} = \psi_l a_{k+1}$

$$+ \sum_{j=0}^{\infty} \psi_{l+1+j}a_{k-j} = \hat{w}_k(l+1) - \psi_l\hat{w}_k(1)$$

$$+ \psi_l w_{k+1}, \qquad l > 0 \tag{7.2.12}$$

根据以上这些性质和公式，便可以建立关于 AR, MA 或 ARMA 序列的各种具体的预报方法，我们将在下节一一介绍.

为了给出 ARIMA 和季节性序列的预报方法，还须要对这类序列的预报的定义和性质加以特别的讨论.

先讨论 ARIMA(p,d,q) 序列的情况. 在第二章 §3 中我们曾给出这种序列 z_t 的定义，它们是从(2.3.3)式即 $\nabla^d z_t = w_t (t > d)$ 解出来的非平稳序列，其中 w_t 是 ARMA (p, q) 序列，z_t 的初始状态值 $z_1, z_2, \cdots, z_d$ 是与 w_t 不相关的随机变量，在我们的讨论中，还假定它们的联合分布是正态分布. 这时，对于 z_{k+l} 的预报 $\hat{z}_k(l)$ 是通过平稳 ARMA 序列 w_t 的平稳预报来定义的，即

$$\hat{z}_k(l) = E(z_{k+l}|z_1, z_2, \cdots, z_d, w_k, w_{k-1}, \cdots) = E(z_{k+l}|\mathscr{Z}_k)$$

其中

$$\mathscr{Z}_k = \left\{z:\ z = w + \sum_{j=1}^{d} C_j z_j, w \in \mathscr{W}_k,\ C_j \text{为任意实数}\right\}$$

当然，在实际问题中我们并不知道 $w_j\ (j \leqslant d)$，我们所以这样定义 $\hat{z}_k(l)$，一方面因为理论分析简单，而另一方面因为当 k 很大以后，可以用有穷历史数据取近似的 $\hat{z}_k(l)$. 根据 (2.3.5) 式可知

$$\hat{z}_k(l) = \sum_{i=0}^{d-1} C_i^{l+i-1}\nabla^i z_k + \sum_{j=1}^{k+l-d} C_{d-1}^{l-j+d-1}E(w_{k+j}|\mathscr{Z}_k)$$

$$= \sum_{i=0}^{d-1} C_i^{l+i-1}\nabla^i z_k + \sum_{j=1}^{l} C_{d-1}^{l-j+d-1}\hat{w}_k(j), \ l > 0 \tag{7.2.13}$$

由于 w_k 与 $z_1, z_2, \cdots, z_{d-1}$ 独立，所以 $E(w_{k+j}|\mathscr{Z}) = E(w_{k+j}|\mathscr{W}_k) = \hat{w}_k(j)$. 由此知 $\hat{z}_k(l)$ 便可用 $\nabla^i z_k (0 \leqslant i < d)$ 以及 $\hat{w}_k(j) (1 \leqslant j \leqslant l)$ 表示出来，只要能计算出这些量，也就得

到 $\hat{z}_k(l)$ 了. 下面以 $d=1$ 和 $d=2$ 两个常用的特殊情况为例，列出它们的预报公式:

$d=1$ 时，由 (2.3.6) 式直接可以得到

$$\hat{z}_k(l)=z_k+\sum_{j=1}^{l}\hat{w}_k(j),\qquad l>0 \tag{7.2.14}$$

$d=2$ 时，由 (2.3.7) 式可以得到

$$\hat{z}_k(l)=z_k+l(z_k-z_{k-1})+\sum_{j=1}^{l}(l+1-j)\hat{w}_k(j)$$
$$l>0 \tag{7.2.15}$$

关于季节性模型，我们仍限于分析两种常用的简单特例，通过它们就可了解这种序列预报的特性，我们讨论 $(1-B^s)z_t$ 和 $(1-B)(1-B^s)z_t$ 为 ARMA 序列时的预报问题.

若 $w_t=(1-B^s)z_t$ 为 ARMA(p,q) 序列，设 r 为满足 $(r-1)s<l\leqslant sr$ 的整数，则

$$\begin{aligned}z_{k+l}&=z_{k+l}-z_{k+l-s}+z_{k+l-s}-\cdots+z_{k+l-rs}\\&=z_{k+l-rs}+\sum_{j=0}^{r-1}w_{k+l-js}\end{aligned} \tag{7.2.16}$$

由此可得

$$\hat{z}_k(l)=z_{k+l-rs}+\sum_{j=0}^{r-1}\hat{w}_k(l-js) \tag{7.2.17}$$

若 $(1-B)(1-B^s)z_t=w_t$ 为 ARMA(p,q)序列，令 $v_t=(1-B^s)z_t$，则 $w_t=(1-B)v_t$，仿照 (7.2.14) 式可知

$$\hat{v}_k(l)=v_k+\sum_{j=1}^{l}\hat{w}_k(j)$$

再令 $(r-1)s<l\leqslant rs$，仿照 (7.2.16) 式有

$$\hat{z}_k(l)=z_{k+l-rs}+\sum_{j=0}^{r-1}\hat{v}_k(l-js)=z_{k+l-rs}$$
$$+\sum_{j=0}^{r-1}\left[v_k+\sum_{i=1}^{l-js}\hat{w}_k(i)\right]=z_{k+l-rs}+r(z_k-z_{k-s})$$

$$+\sum_{j=1}^{l}\left(\left[\frac{l-j}{s}\right]+1\right)\hat{w}_k(j), \quad l>0 \qquad (7.2.18)$$

其中 $\left[\frac{l-j}{s}\right]$ 表 $\frac{l-j}{s}$ 的整数部分.

以上诸式中的 $\hat{w}_k(j)$ 都是在 $\hat{w}_k(j)=E(w_{k+j}|\mathscr{W}_k)$ 意义下的预报值，只有这样，才能借助于平稳序列的平稳预报，给出 z_{k+l} 的预报值 $\hat{z}_k(l)$ 的简单表达式.当然，由于 z_t 本身为非平稳序列，虽然用平稳预报方法定义 $\hat{z}_k(l)$，但当固定 l 时，$\hat{z}_k(l)$，$k=1,2,\cdots$不是平稳序列. 为了区别于§4中介绍的适时预报方法，我们仍统称本节给出的所有预报为平稳预报.

§3. 各类序列的预报方法

一、AR 序列的预报方法

1. 递推方法，即根据(7.2.1)式（此时 $q=0$）可得

$$\hat{w}_k(l)=\varphi_1\hat{w}_k(l-1)+\varphi_2\hat{w}_k(l-2)+\cdots+\varphi_p\hat{w}_k(l-p)$$
$$l>0 \qquad (7.3.1)$$

又由 $\hat{w}_k(-l)=w_{k-l}\ (l\geqslant 0)$ 便可得上式的初始值. (7.3.1)式给出了 AR 序列预报的递推公式，必须注意，它是关于 l 的递推公式.

2. 预报函数方法：即给出(7.3.1)式的解

$$\hat{w}_k(l)=C_k(l)=\sum_{i=1}^{p}b_i^{(k)}f_i(l), \quad l>-p \qquad (7.3.2)$$

初始值仍为 $\hat{w}_k(-j)=w_{k-j}\ (j=0,1,\cdots,p-1)$，即 $b_i^{(k)}$ 满足

$$\sum_{i=1}^{p}b_i^{(k)}f_i(-j)=w_{k-j}, \quad j=0,1,\cdots,p-1 \qquad (7.3.3)$$

由此可确定(7.3.2)式中的系数 $b_i^{(k)}$，其中 $f_i(l)$ 是由 $\varphi(B)=0$ 的根所决定的 l 的指数、三角函数或幂函数之积. 这已经在第七章§1里指出过了.

由此可见，对于 AR$(p,0)$ 序列而言，其预报值 $\hat{w}_k(l)(l\geqslant 0)$

只须用到 $w_k, w_{k-1}, \cdots, w_{k-p+1}$，而用不着更早的历史值 $w_j(j \leqslant k-p)$. 因此，当 $k \geqslant p$ 时，便可严格地算出 $\hat{w}_k(l)$ 之值，而无须用近似处理方法.

例 1. $(1-\varphi_1 B)w_t = a_t$ 的预报. 若用递推法，由 (7.3.1) 式可得

$$\hat{w}_k(l+1) = \varphi_1 \hat{w}_k(l), \quad l > 0$$

初始值 $\hat{w}_k(0) = w_k$. 若用预报函数法，由 (7.3.2) 式可知

$$\hat{w}_k(l) = b_1^{(k)} \varphi_1^l$$

由于 $w_k = \hat{w}_k(0) = b_1^{(k)} \varphi_1^0 = b_1^{(k)}$，所以

$$\hat{w}_k(l) = w_k \varphi_1^l, \quad l \geqslant 0$$

例 2. $(1-\varphi_1 B-\varphi_2 B^2)w_t = a_t$ 的预报. 根据(7.3.1)式可得差分方程

$$\hat{w}_k(l) = \varphi_1 \hat{w}_k(l-1) + \varphi_2 \hat{w}_k(l-2), \ l > 0$$

初始值 $\hat{w}_k(0) = w_k$, $\hat{w}_k(-1) = w_{k-1}$. 令

$$\varphi(B) = 1-\varphi_1 B-\varphi_2 B^2 = (1-G_1 B)(1-G_2 B)$$

即 G_1^{-1}, G_2^{-1} 为 $\varphi(\omega) = 0$ 的两个根. 当 G_1, G_2 为两个不同的实数时，由 (7.3.2) 式，此差分方程的解即预报函数为

$$\hat{w}_k(l) = C_k(l) = b_1^{(k)} G_1^l + b_2^{(k)} G_2^l, \quad l > -2$$

其中 $b_1^{(k)}$, $b_2^{(k)}$ 满足方程

$$\begin{cases} b_1^{(k)} + b_2^{(k)} = w_k \\ b_1^{(k)} G_1^{-1} + b_2^{(k)} G_2^{-1} = w_{k-1} \end{cases}$$

由此解出

$$b_1^{(k)} = \frac{w_k - G_2 w_{k-1}}{1-G_2 G_1^{-1}}, \qquad b_2^{(k)} = \frac{w_k - G_1 w_{k-1}}{1-G_1 G_2^{-1}}$$

$$\hat{w}_k(l) = w_k \frac{G_1^{l+1} - G_2^{l+1}}{G_1 - G_2} - w_{k-1} \frac{G_1 G_2 (G_1^l - G_2^l)}{G_1 - G_2}$$

当 $G_1 = G_2$ 时（此时 G_1 必为实数），预报函数为

$$\hat{w}_k(l) = C_k(l) = b_1^{(k)} G_1^l + l b_2^{(k)} G_1^l, \quad l > -2$$

$b_1^{(k)}$, $b_2^{(k)}$ 满足方程

$$\begin{cases} b_1^{(k)} = w_k \\ (b_1^{(k)} - b_2^{(k)})G_1^{-1} = w_{k-1} \end{cases}$$

其解为 $b_1^{(k)} = w_k$，$b_2^{(k)} = w_k - G_1 w_{k-1}$，

$$\hat{w}_k(l) = (l+1)w_k G_1^l - l w_{k-1} G_1^{l+1}$$

当 G_1 为复数 $\rho(\cos\gamma + i\sin\gamma)$ 时，必有 $G_2 = \bar{G}_1 = \rho(\cos\gamma - i\sin\gamma)$．这时易知预报函数可表为

$$\hat{w}_k(l) = C_k(l) = b_1^{(k)}\rho^l \cos l\gamma + b_2^{(k)}\rho^l \sin l\gamma, \quad l > -2$$

其中系数满足方程

$$\begin{cases} b_1^{(k)} = w_k \\ b_1^{(k)}\rho^{-1}\cos\gamma - b_2^{(k)}\rho^{-1}\sin\gamma = w_{k-1} \end{cases}$$

其解为 $b_1^{(k)} = w_k$，$\quad b_2^{(k)} = \dfrac{w_k\cos\gamma - \rho w_{k-1}}{\sin\gamma}$，

$$\hat{w}_k(l) = w_k\rho^l \frac{\sin(l+1)\gamma}{\sin\gamma} - w_{k-1}\rho^{l+1}\frac{\sin l\gamma}{\sin\gamma}$$

二、MA 序列的预报方法

1. 直接方法：对于 MA$(0, q)$ 序列的预报，由(7.2.4)和(7.2.2)式可知，

$$\hat{w}_k(l) = \begin{cases} \sum_{j=1}^{\infty} \pi_j^{(l)} w_{k+1-j}, & 0 < l \leqslant q \\ 0, & l > q \end{cases} \tag{7.3.4}$$

由此可见，它与 AR 序列预报的一个明显差别是，(7.3.4)式中的 $\hat{w}_k(l)(0 < l \leqslant q)$ 要用到全部历史的数据 $w_{k-j}(j \leqslant 0)$．但是真实数据总是有限的，所以实际应用只能由有穷和来近似，即取

$$\hat{w}_k(l) \simeq \sum_{j=1}^{k} \pi_j^{(l)} w_{k+1-j}, \quad 0 < l \leqslant q \tag{7.3.5}$$

我们已经指出过，对于平稳可逆的 ARMA 序列，有 $|\pi_j| \leqslant d_1 e^{-d_2|j|}$，于是由(7.2.5)式不难证明

$$|\pi_j^{(l)}| \leqslant d_1(l) e^{-d_2(l)|j|}$$

此处 $d_1(l)$ 与 $d_2(l)$ 是与 l 有关的常数．因此，当 k 很大时，(7.3.5)

式是 (7.3.4) 式的很好的近似；甚至 (7.3.5) 式中的求和只须从 1 加到某个适当选取的 $k_0(k_0 < k)$，也能保证一定的近似精度.

对于 AR(p, 0) 序列而言，为要给出各个未来时刻 $w_{k+l}(l > 0)$的预报值 $\hat{w}_k(l)$，既可以由 (7.3.1) 式逐个递推计算，也可以由 (7.3.1) 式的具体解式 (7.3.2) 直接给出. 对于 MA(0, q) 序列而言，当 $l > q$ 以后的 $\hat{w}_k(l) = 0$，所以为要掌握对各个未来时刻的预报值，实际上只须知道 $\hat{w}_k(l)$ $(l = 1, 2, \cdots, q)$ 就够了. 从表面上看来，似乎比 AR 序列更简单. 但由 (7.3.5) 式可以看到，在每次获得新的数据以后，为了求出 MA(0, q) 序列的 $\hat{w}_k(l)$ $(0 < l \leqslant q)$ 的近似值必须与大量的历史数据一道重新计算式中的加权和. 而对 AR(p, 0) 序列，只须通过计算简单的 (7.3.2) 式得到，而它仅用到很少的 p 个历史数据.

在不少实际应用中，预报是要求连续进行的，比如用电负荷预报、地球自转速率变化预报等，都要每日每时地进行. 同时，每获得一个新的量测（或记录）数据 w_k，就要根据历史数据 w_k, w_{k-1}, $\cdots$, w_1 对未来时刻重新进行一次预报，而 k 是随时间推移不断增加的. 特别，如果采样间隔的时间很短（比如几分钟或几秒钟），就会感到用 (7.3.5) 式直接求 $\hat{w}_k(l)$ 的计算量太大了，甚至可能在采样间隔时间以内来不及计算完各个预报值 $\hat{w}_k(l)$；并且数据的存贮量也越来越大. 对于应用者来说，就迫切希望能用关于 k 的递推公式计算 $\hat{w}_k(l)$，以减少计算量和存贮量. 以上所述，对于 AR(p, 0) 序列的预报，这个困难并不存在，因为每次求 $\hat{w}_k(l)$ 的计算量不大，而且只须保留 p 个历史数据. 为了适应这类适时预报的需要，我们在关于 MA 和 ARMA 序列的讨论中，都特别介绍一种关于 k 的递推预报方法. 但对于那些非连续性预报，或者虽然连续，但采样间隔较长（如一小时或一天等）的实际问题，那就没有必要采用这种递推格式.

2. 预报向量递推方法：如上所述，在对 MA(0, q) 序列进行预报时，需要从 $\hat{w}_k(l)(0 < l \leqslant q)$ 和新获得的数据 w_{k+1}，通过递推公式求出 $\hat{w}_{k+1}(l)$ $(0 < l \leqslant q)$. 我们令

$$\hat{\boldsymbol{\omega}}_k^q=(\hat{\omega}_k(1),\hat{\omega}_k(2),\cdots,\hat{\omega}_k(q))^{\tau}$$

并称 $\hat{\boldsymbol{\omega}}_k^q$ 为 MA(0, q) 序列的预报向量. 它和 $\hat{\omega}_k(l)=0\ (l>q)$ 一起描述了在 k 时刻对未来的全部预报结果. 由于对 MA(0, q) 序列, $\psi_k=-\theta_k\ (1\leqslant k\leqslant q)$, 应用 (7.2.12) 式 (取 $l=1,2,\cdots,q$) 即可知

$$\hat{\boldsymbol{\omega}}_{k+1}^q=\begin{pmatrix}\theta_1 & 1 & & 0\\ \theta_2 & & 1 & \\ \vdots & & & \ddots \\ \vdots & 0 & & & 1\\ \theta_q & 0 & \cdots\cdots & & 0\end{pmatrix}\hat{\boldsymbol{\omega}}_k^q-\begin{pmatrix}\theta_1\\ \theta_2\\ \vdots\\ \vdots\\ \theta_q\end{pmatrix}w_{k+1} \qquad (7.3.6)$$

用 (7.3.6) 式便可以递推计算 $\hat{\omega}_k^q$. 这样, 对每一时刻 k, 求 $\hat{\boldsymbol{\omega}}_k^q$ 的计算量都是相同的. 至于递推的初始值可以取某个时刻 k_0 的 $\hat{\boldsymbol{\omega}}_{k_0}^q$, 它可以用直接法求出; 若 k_0 很小时, 也可取做零. 由于 (7.3.6) 式的迭代格式是稳定的, 当 k 增长到很大以后, 初始值的误差影响会逐渐消失.

例3. $w_t=a_t-\theta_1 a_{t-1}$ 的预报. 注意 $\pi_j^{(1)}=\pi_j=-\theta_1^j\ (j\geqslant 1)$, 由直接法有

$$\hat{\omega}_k(1)=\sum_{j=1}^{\infty}\pi_j w_{k+1-j}=-\theta_1\sum_{i=0}^{\infty}\theta_1^i w_{k-i}\cong-\theta_1\sum_{i=0}^{k_0}\theta_1^i w_{k-i}$$

其中取 $k_0\leqslant k$, 以使 $\sum\limits_{i=k_0+1}^{\infty}\theta_1^i$ 可以被忽略 (例如比所要求的相对误差精度还小) 为准. 当 $|\theta_1|$ 接近 1 时, 上述近似公式需要 k_0 很大. 可见用直接法求 $\hat{\omega}_k(1)$ 的计算量至少大于 k_0 个求积. 若用向量递推式 (7.3.6) 式, 则有

$$\hat{\omega}_{k+1}(1)=\theta_1\hat{\omega}_k(1)-\theta_1\hat{\omega}_{k+1}$$

在连续进行预报时, 用此式计算 $\hat{\omega}_{k+1}(1)$, 只需两个乘法与一个减法运算. 当 k 增长到很大以后, $\hat{\omega}_k(1)$ 同样有很好的近似精度.

例 4. $w_t=a_t-a_{t-1}+0.24a_{t-2}$ 的预报. 注意

$$1-B+0.24B^2=(1-0.4B)(1-0.6B)$$

于是

$$1-\sum_{j=1}^{\infty}\pi_j B^j=\theta^{-1}(B)=\frac{1}{(1-0.4B)(1-0.6B)}$$

$$=\frac{3}{1-0.6B}-\frac{2}{1-0.4B}$$

$$=\sum_{j=0}^{\infty}[3(0.6)^j-2(0.4)^j]B^j$$

即 $\pi_j=2(0.4)^j-3(0.6)^j\ (j\geqslant 1)$，因而由 (7.2.5) 式，

$$\pi_j^{(2)}=\pi_{j+1}+\pi_1\pi_j=1.2[(0.6)^j-(0.4)^j]\quad (j\geqslant 1)$$

容易验证，当 $k\geqslant 13$ 时，$\sum_{j\geqslant k}|\pi_j|$ 和 $\sum_{j\geqslant k}|\pi_j^{(2)}|$ 都小于 0.01．若我们要求的相对误差精度为 1/10，则 0.01 便可以忽略，于是可以选定 $k_0=13$．由直接法可知

$$\hat{w}_k(1)=\sum_{j=1}^{\infty}\pi_j w_{k+1-j}\cong\sum_{j=1}^{13}[2(0.4)^j-3(0.6)^j]w_{k+1-j}$$

$$\hat{w}_k(2)=\sum_{j=1}^{\infty}\pi_j^{(2)}w_{k+1-j}\cong 1.2\sum_{j=1}^{13}[(0.6)^j-(0.4)^j]w_{k+1-j}$$

若用预报向量方法，由 (7.3.6) 式即有

$$\hat{w}_{k+1}^2=\begin{pmatrix}\theta_1 & 1\\ \theta_2 & 0\end{pmatrix}\hat{w}_k^2-\begin{pmatrix}\theta_1\\ \theta_2\end{pmatrix}w_{k+1}$$

$$=\begin{pmatrix}1 & 1\\ -0.24 & 0\end{pmatrix}\hat{w}_k^2+\begin{pmatrix}-1\\ 0.24\end{pmatrix}w_{k+1}$$

三、ARMA 序列的预报方法

1. 直接方法：与 MA 序列预报的情形一样，由 (7.2.4) 式计算 $\hat{w}_k(l)$ 便是直接方法．所不同的是在 ARMA 序列情况下，$\hat{w}_k(\theta)$ 在 $l>q$ 以后不全为零．

2. 预报向量递推方法：它和 MA 序列预报向量递推方法相类似，但是稍微复杂一些．在介绍了 MA 序列预报的直接方法之后，我们曾指出，为了掌握 MA 序列的全部未来时刻预报值，只须掌握预报向量值 $\hat{w}_k^q$（因为 $\hat{w}_k(l)=0,\ l>q$）；为了掌握 AR 序列的

全部预报值，只须给出其预报函数 $C_k(l)$ 的具体表达式. 对于 ARMA(p,q) 序列而言，从(7.2.7)式可知，当 $l>q$ 时，为了知道以后的全部 $\hat{w}_k(l)$，只要由 $\hat{w}_k(q)$, $\hat{w}_k(q-1)$, $\cdots$, $\hat{w}_k(q-p+1)$ 定出它的预报函数具体表达式的诸系数 $b_i^{(k)}(1\leqslant i\leqslant p)$ 即可；而当 $0<l\leqslant q$ 时，$\hat{w}_k(l)$ 一般不便用预报函数计算. 换句话说，为了给出 ARMA 序列的全部未来时刻的预报值，我们必须且只须算出 $\hat{w}_k(1)$, $\hat{w}_k(2)$, $\cdots$, $\hat{w}_k(q)$. 为此，我们仍定义

$$\hat{\boldsymbol{w}}_k^q=(\hat{w}_k(1),\hat{w}_k(2),\cdots,\hat{w}_k(q))^\tau$$

为 ARMA 序列的预报向量. 于是，应用(7.2.12)式(取 $l=1,2,\cdots,q$)和(26.1)式(取 $l=q+1$)可得

$$\hat{\boldsymbol{w}}_{k+1}^q=\begin{pmatrix}-\psi_1 & 1 & 0 & \cdots\cdots & 0 & 0\\ -\psi_2 & 0 & 1 & \cdots\cdots & 0 & 0\\ & \cdots\cdots & & & & \\ -\psi_{q-1} & 0 & 0 & \cdots\cdots & 0 & 1\\ -\psi_q+\varphi_q' & \varphi_{q-1}' & \varphi_{q-2}' & \cdots & \varphi_2' & \varphi_1'\end{pmatrix}\hat{\boldsymbol{w}}_k^q$$

$$+\begin{pmatrix}\psi_1\\ \psi_2\\ \vdots\\ \psi_{q-1}\\ \psi_q\end{pmatrix}w_{k+1}+\begin{pmatrix}0\\ 0\\ \vdots\\ 0\\ \sum\limits_{j=q+1}^{p}\varphi_j' w_{k+q-j+1}\end{pmatrix}\tag{7.3.7}$$

其中 $\varphi_j'=\varphi_j(0<j\leqslant p)$; $\varphi_j'=0\ (j>p)$，而且当 $p\leqslant q$ 时，上式第三项为零. 至于(7.3.7)式的初始值 $\hat{w}_{k_0}^q$ 的取法，与 MA 序列的情况完全相同，可由直接法求得，也可用零值代替(当 k_0 很小时).

例 5. $(1-0.8B)w_t=(1-B+0.24B^2)a_t$ 的预报. 将原模型写作

$$w_t=(1-0.8B)^{-1}(1-B+0.24B^2)a_t$$

$$=\sum_{i=0}^{\infty}(1-B+0.24B^2)(0.8B)^i a_t$$

由此易见，$|\phi_j|$ 与 $(0.8)^j$ 为相同数量级. 因此，为了保证有限和的误差精度为 0.01，只要在(7.2.4)式中取 k_0 为 30 左右即可，由此便可得直接方法的近似预报公式.

为了采用递推方法，令 $\hat{\boldsymbol{\omega}}_k^2=(\hat{\omega}_k(1),\ \hat{\omega}_k(2))^\tau$，由(7.3.7)式可知

$$\hat{\boldsymbol{\omega}}_{k+1}^2=\begin{pmatrix}-\phi_1 & 1\\ -\phi_2 & \varphi_1\end{pmatrix}\hat{\boldsymbol{\omega}}_k^2+\begin{pmatrix}\phi_1\\ \phi_2\end{pmatrix}w_{k+1}$$

$$=\begin{pmatrix}0.2 & 1\\ -0.08 & 0.8\end{pmatrix}\hat{\boldsymbol{\omega}}_k^2+\begin{pmatrix}-0.2\\ 0.08\end{pmatrix}w_{k+1}$$

为了给出其余的 $\hat{\omega}_k(l)\ (l>2)$，由 (7.2.7) 式知

$$\hat{\omega}_k(l)=C_k(l)=b_1^{(k)}\varphi_1^l=b_1^{(k)}(0.8)^l,\quad l>1$$

又由 (7.2.8) 式可知

$$b_1^{(k)}(0.8)^2=\hat{\omega}_k(2)$$

即 $b_1^{(k)}=\dfrac{1}{0.64}\hat{\omega}_k(2)$，于是

$$\hat{\omega}_k(l)=\hat{\omega}_k(2)(0.8)^{l-2},\quad l\geqslant 2$$

四、ARIMA 序列的预报方法

无论 ARIMA 序列，还是满足季节性模型的序列，都是通过对 ARMA 型平稳序列的求和运算得到的一种非平稳序列. 因此，根据 §2 末尾的叙述，这类序列的预报是通过平稳序列的平稳预报方法来定义和计算的. 即根据产生这些非平稳序列的 $w_t=\nabla^d z_t$ (或 $w_t=\nabla^d\nabla_s^D z_t$) 的预报值 $\hat{\omega}_k(l)$，求出这些序列本身的预报值 $\hat{z}_k(l)$ (如 (7.2.13)，(7.2.18) 等式).

下面介绍计算 ARIMA 序列预报的方法. 设 z_t 满足以下的季节性模型

$$\nabla^d\nabla_s^D\varphi(B)z_t=\theta(B)a_t \tag{7.3.8}$$

当 $D=0$ 时，便是普通的 ARIMA (p,d,q) 模型，当 $D>0$ 时，便为季节性模型. 记 $w_t=\nabla^d\nabla_s^D z_t$，则 w_t 为 ARMA (p,q)序列，因此，$\hat{\omega}_k(l)$ 满足预报差分方程 (7.2.1) 式，而 (7.2.7) 式是此方

程的解．现在记 $\varphi(B)=\nabla^d\nabla_s^D\varphi(B)$，由于估计量 $\hat{\omega}_k(l)$ 具有线性性质，所以 $\hat{\omega}_k(l)=\nabla^d\nabla_s^D\hat{z}_k(l)$ $(k>d+D, l\geqslant 0)$，其中 k 固定，算子 ∇，∇_s 都作用在 l 上．于是将此式代入(7.2.1)式便得到（当 $k>d+D$ 时）

$$\varphi(B)\hat{z}_k(l)=0, \qquad l>q \tag{7.3.9}$$

这个差分方程的通解形式为

$$\hat{z}_k(l)=\sum_{i=1}^{p+d+Ds}b_i^{(k)}f_i(l), \quad l>q-(p+d+Ds) \tag{7.3.10}$$

其中 $f_i(l)$ $(1\leqslant i\leqslant p+d+Ds)$ 是(7.3.9)式的 $p+d+Ds$ 个线性无关解，它们可表为 l 的指数函数、幂函数和三角函数的乘积形式，$b_i^{(k)}$ 可由初值 $\hat{z}_k(q)$，$\hat{z}_k(q-1)$，$\cdots$，$\hat{z}_k(q-p-d-Ds+1)$ 确定，我们把(7.3.10)式称为 z_t 的预报函数．

特别，如果(7.3.8)式中的 $D=0$，则 $\varphi(\omega)=0$ 在单位圆周上仅以 1 为根，而且是 d 重根．这时(7.3.10)式可以写成

$$\hat{z}_k(l)=h_0^{(k)}+h_1^{(k)}l+\cdots+h_{d-1}^{(k)}l^{d-1}+\sum_{i=1}^{p}b_i^{(k)}f_i(l)$$

$$l>q-(p+d) \tag{7.3.11}$$

更特别，如果(7.3.8)式中的 $p=0$，$D=0$ 则 $\hat{z}_k(l)$ 只剩下(7.3.11)式中的多项式部分．

至于与 ARMA 序列预报相应的直接计算 $\hat{z}_k(l)$ 方法，以及递推计算法这里不一一叙述了．下面通过几种常用的低阶求和序列和季节性序列的预报的例子，以熟悉这类序列的预报方法．

例 6. $(1-B)z_t=(1-\theta_1 B)a_t$ 的预报．由(7.2.14)式和(7.2.2)式立刻可知

$$\hat{z}_k(l)=z_k+\sum_{j=1}^{l}\hat{\omega}_k(j)=z_k+\hat{\omega}_k(1)=\hat{z}_k(1), \quad l\geqslant 1 \tag{7.3.12}$$

其中 $\hat{\omega}_k(1)$ 可以用直接方法计算（见例3），于是

$$\hat{z}_k(1) = z_k - \theta_1 \sum_{j=0}^{\infty} \theta_1^j w_{k-j} \cong z_k - \theta_1 \sum_{j=1}^{k-2} \theta_1^j (z_{k-j} - z_{k-j-1})$$

(7.3.13)

同样地，也可以由 $\hat{w}_k(1)$ 的递推公式（见例3）得到

$$\begin{aligned}\hat{z}_{k+1}(1) &= z_{k+1} + \hat{w}_{k+1}(1) = z_{k+1} + \theta_1 \hat{w}_k(1) - \theta_1 w_{k+1} \\ &= z_{k+1} + \theta_1[\hat{w}_k(1) - z_{k+1} + z_k] \\ &= (1-\theta_1) z_{k+1} + \theta_1 \hat{z}_k(1) \qquad (7.3.14)\end{aligned}$$

这就是 $\hat{z}_k(1)$ 的递推式．初值可以由直接法计算．

还可以用 z_t 的预报函数来求 $\hat{z}_k(l)$ $(l>1)$．由于此时 $p=0$，$d=1$，所以由(7.3.11)式，$\hat{z}_k(l) = h_0^{(k)} = \hat{z}_k(1)$ $(l>1)$．这与(7.3.12)式相同．

例 7. $(1-\varphi_1 B)(1-B)z_t = a_t$ 的预报．记 $w_t = z_t - z_{t-1}$，由例 1 知 $\hat{w}_k(j) = w_k \varphi_1^j$ $(j \geqslant 0)$，代入(7.2.4)式得

$$\begin{aligned}\hat{z}_k(l) &= z_k + \sum_{j=1}^{l} w_k \varphi_1^j = z_k + \frac{\varphi_1(1-\varphi_1^l)}{1-\varphi_1}(z_k - z_{k-1}) \\ &= \frac{1-\varphi_1^{l+1}}{1-\varphi_1} z_k - \frac{\varphi_1(1-\varphi_1^l)}{1-\varphi_1} z_{k-1} \qquad (7.3.15)\end{aligned}$$

而 z_t 的预报函数为

$$\hat{z}_k(l) = h_0^{(k)} + h_1^{(k)} \varphi_1^l, \quad l > -2 \qquad (7.3.16)$$

由 $z_k = \hat{z}_k(0) = h_0^{(k)} + h_1^{(k)}$，$z_{k-1} = \hat{z}_k(-1) = h_0^{(k)} + h_1^{(k)} \varphi_1^{-1}$ 可解出

$$h_0^{(k)} = \frac{z_k - \varphi_1 z_{k-1}}{1-\varphi_1}, \qquad h_1^{(k)} = \frac{\varphi_1}{1-\varphi_1}(z_{k-1} - z_k)$$

当然，把它们代入(7.3.16)式后，所得的就是(7.3.15)式．此时由于直接计算 $\hat{z}_k(l)$ 的公式十分简单，不必使用关于 k 的递推公式．

例 8. 在第六章末尾引用过地球自转速率变化过程的例子，我们根据历史观测数据，用那里介绍的方法，分析确定了它的估计模型如下：

$$\nabla \nabla_{73}(1 - 0.956B) z_t = a_t$$

现在根据这个模型来求 z_t 的预报值 $\hat{z}_k(l)$．我们仅考虑未来时刻

在一年以内，即 $(1 \leqslant l \leqslant 73)$ 的预报情况. 应用 (7.2.18) 式，其中 $r=1$，可知

$$\hat{z}_k(l)=z_{k+l-73}+z_k-z_{k-73}+\sum_{j=1}^{l}\hat{w}_k(j)$$

注意此时 $w_t=\nabla\nabla_{73}z_t$ 为 AR(1, 0) 序列，所以进一步有

$$\begin{aligned}\hat{z}_k(l)&=z_{k+l-73}+z_k-z_{k-73}+\sum_{j=1}^{l}w_k\varphi_1^j\\&=z_{k+l-73}+z_k-z_{k-73}+\frac{\varphi_1(1-\varphi_1^l)}{1-\varphi_1}w_k\\&=z_{k+l-73}+z_k-z_{k-73}+\frac{0.956[1-(0.956)^l]}{0.044}\\&\quad\times(z_k-z_{k-1}-z_{k-73}+z_{k-74})\end{aligned}\tag{7.3.17}$$

在表 7.3.1 中列出了根据 1967 年 3 月 25 日以前的数据按上式预报方法，对于半年以内（每五天报一个值）的未来值的预报和误差. 就这些预报结果看来，还是比较合乎实际情况的.

表 7.3.1　地球自转速率预报结果(误差 $=\hat{z}_k(l)-z_{k+l}$)

l	$\hat{z}_k(l)$	误差	l	$\hat{z}_k(l)$	误差	l	$\hat{z}_k(l)$	误差
1	−54663	2	13	−56295	3	25	−57387	−4
2	−54800	1	14	−56410	5	26	−57474	0
3	−54939	2	15	−56519	6	27	−57566	7
4	−55080	6	16	−56624	7	28	−57663	15
5	−55219	4	17	−56722	7	29	−57762	23
6	−55360	2	18	−56809	−1	30	−57865	30
7	−55501	0	19	−56991	−9	31	−57972	41
8	−55642	−3	20	−56980	−7	32	−58083	56
9	−55782	−3	21	−57062	−8	33	−58201	78
10	−55916	−6	22	−57141	−11	34	−58322	101
11	−56047	−4	23	−57222	−10	35	−58447	122
12	−56173	0	24	−57304	−6	36	−58574	143

例9. 我们曾经讨论过用电负荷的例子，按照第六章的方法，在对历史数据资料进行时序分析后，确定其可供拟合的随机模型之一为（见 (6.3.13) 式）

$$(1-\varphi_1 B)(1-\Phi_1 B^{24})\nabla_{24} z_t = a_t, \quad \Phi_1 < 0$$

这时记 $w_t = \nabla_{24} z_t = z_t - z_{t-24}$，则 w_t 满足以下模型:

$$(1-\varphi_1 B)(1-\Phi_1 B^{24}) w_t = a_t$$

于是由 (7.2.17) 式可知，当 $24(r-1) < l \leqslant 24r$ 时，

$$\hat{z}_k(l) = z_{k+l-24r} + \sum_{j=0}^{r-1} \hat{w}_k(l-24j)$$

其中 $\hat{w}_k(l)$ 可按 AR 序列的预报方法求得.

若采用预报函数方法，注意 w_t 的预报差分方程为

$$\hat{w}_k(l) - \varphi_1 \hat{w}_k(l-1) - \Phi_1 \hat{w}_k(l-24) + \varphi_1 \Phi_1 \hat{w}_k(l-25) = 0, \quad l > 0$$

它的通解与 (6.3.14) 式的解的形式完全一样，即

$$\hat{w}_k(l) = b_0^{(k)} \varphi_1^l + |\Phi_1|^{\frac{l}{24}} \sum_{j=1}^{12} \left(b_j'^{(k)} \cos \frac{l(2j+1)\pi}{24} + b_j''^{(k)} \sin \frac{l(2j+1)\pi}{24} \right)$$

可以看出，若记

$$f_l^{(k)} = \sum_{j=1}^{12} \left(b_j'^{(k)} \cos \frac{l(2j+1)\pi}{24} + b_j''^{(k)} \sin \frac{l(2j+1)\pi}{24} \right)$$

则它是 l 的周期长度为 24 的周期序列. 对于任意 $l > 0$，仍设 r 为满足 $24(r-1) < l \leqslant 24r$ 的整数，又设 $l^* = l - 24(r-1)$，则 $1 \leqslant l^* \leqslant 24$，而且有

$$\hat{w}_k(l) = b_0^{(k)} \varphi_1^l + |\Phi_1|^{\frac{l}{24}} f_{l^*}^{(k)}, \qquad l > -25$$

取 $l = 0, -1, -2, \cdots, -24$，则

$$\begin{cases} f_{24}^{(k)} + b_0^{(k)} = \hat{w}_k(0) = w_k = z_k - z_{k-24} \\ |\Phi_1|^{-\frac{1}{24}} f_{23}^{(k)} + \varphi_1^{-1} b_0^{(k)} = \hat{w}_k(-1) = w_{k-1} = z_{k-1} - z_{k-25} \\ \cdots\cdots \\ |\Phi_1|^{-\frac{23}{24}} f_1^{(k)} + \varphi_1^{-23} b_0^{(k)} = \hat{w}_k(-23) = w_{k-23} = z_{k-23} - z_{k-47} \\ |\Phi_1|^{-1} f_{24}^{(k)} + \varphi_1^{-24} b_0^{(k)} = \hat{w}_k(-24) = w_{k-24} = z_{k-24} - z_{k-48} \end{cases}$$

由这个方程组可解出 $b_0^{(k)}$ 和 $f_i^{(k)}$ $(1 \leqslant i \leqslant 24)$，而不必去求 $b_i'^{(k)}$

和 $b_i'^{(k)}(1\leqslant i\leqslant 12)$. 有了 $b_0^{(k)}$ 和 $f_i^{(k)}(1\leqslant i\leqslant 24)$ 以后，再代入 $\hat{z}_k(l)$ 的表达式（即 (7.2.17) 式）就得到

$$\begin{aligned}
\hat{z}_k(l) &= z_{k+l-24r}+\sum_{j=0}^{r-1}\hat{\omega}_k(l-24j)\\
&= z_{k+l-24r}+\sum_{j=0}^{r-1}\left\{b_0^{(k)}\varphi_1^{l-24j}+|\Phi_1|^{\frac{l-24j}{24}}f_{l^*}^{(k)}\right\}\\
&= z_{k+l-24r}+b_0^{(k)}\varphi_1^l\sum_{j=0}^{r-1}\varphi_1^{-24j}+|\Phi_1|^{\frac{l}{24}}f_{l^*}^{(k)}\sum_{j=0}^{r-1}|\Phi_1|^{-j}\\
&= z_{k+l-24r}+b_0^{(k)}\frac{\varphi_1^l-\varphi_1^{l-24r}}{1-\varphi_1^{-24}}\\
&\quad+f_{l^*}^{(k)}\frac{|\Phi_1|^{\frac{l}{24}}-|\Phi_1|^{\frac{l-24r}{24}}}{1-|\Phi_1|^{-1}},\qquad l>0
\end{aligned}$$

这是计算 $\hat{z}_k(l)$ 的直接公式，其中 $b_0^{(k)}$ 和 $f_{l^*}^{(k)}$ $(1\leqslant l^*\leqslant 24)$ 也直接由 z_{k-j} $(j\geqslant 0)$ 算出.

例 10. 仍以电力负荷为例，现在考虑采用另一拟合模型（见 (6.3.11) 式）

$$\nabla\nabla_{24}z_t=(1-\theta_1B)(1-\Theta_1B^{24})a_t$$

来讨论它的预报问题. 这时，令 $w_t=\nabla\nabla_{24}z_t=z_t-z_{t-1}-z_{t-24}+z_{t-25}$，它满足以下模型:

$$w_t=a_t-\theta_1a_{t-1}-\Theta_1a_{t-24}+\theta_1\Theta_1a_{t-25}$$

于是，$\hat{w}_k(l)$ 可由 MA 序列的预报方法进行预报，而 $\hat{z}_k(l)$ 由 (7.2.18) 式间接得到. 为此先求出 w_t 的逆转算子形式，即

$$\begin{aligned}
1-\sum_{j=1}^{\infty}\pi_jB^j &= (1-\theta_1B)^{-1}(1-\Theta_1B^{24})^{-1}\\
&= \left(\sum_{j=0}^{\infty}\theta_1^jB^j\right)\left(\sum_{j=0}^{\infty}\Theta_1^jB^{24j}\right)\\
&= \sum_{j=0}^{\infty}\left\{\theta_1^j\frac{1-(\theta_1^{-24}\Theta_1)^{\left[\frac{j}{24}\right]+1}}{1-\theta_1^{-24}\Theta_1}\right\}B^j
\end{aligned}$$

其中 $\left[\frac{i}{24}\right]$ 表示 $\frac{i}{24}$ 的整数部分. 于是

$$x_i = -\theta_1^i \frac{1-(\theta_1^{-24}\Theta_1)^{\left[\frac{i}{24}\right]+1}}{1-\theta_1^{-24}\Theta_1} = -\theta_1^i \frac{\theta_1^{24}-\Theta_1^{\left[\frac{i}{24}\right]+1}\theta_1^{-24\left[\frac{i}{24}\right]}}{\theta_1^{24}-\Theta_1}$$

$$\cong -\theta_1^{i-24\left[\frac{i}{24}\right]}\Theta_1^{\left[\frac{i}{24}\right]}$$

其中 $\theta_1^{24} \cong 0$. 将此式代入 (7.2.5) 和 (7.3.5) 式便可得到 $\hat{w}_k(l)$ $(1 \leqslant l \leqslant 25)$ 的近似估计值.

为要使用递推方法进行连续预报,我们引入记号

$$F = \begin{pmatrix} \theta_1 & 1 & & & 0 & \\ 0 & 0 & \ddots & & & \\ \vdots & \vdots & & \ddots & \ddots & \\ 0 & & & & \ddots & \ddots \\ \Theta_1 & 0 & \cdots & \cdots & 0 & 1 \\ -\theta_1\Theta_1 & 0 & \cdots & \cdots & 0 & 0 \end{pmatrix}_{25\times 25} \qquad G = \begin{pmatrix} -\theta_1 \\ 0 \\ \vdots \\ 0 \\ -\Theta_1 \\ \theta_1\Theta_1 \end{pmatrix}_{25\times 1}$$

$$\hat{\boldsymbol{w}}_k^{25} = (\hat{w}_k(1), \hat{w}_k(2), \cdots, \hat{w}_k(25))^{\tau}$$

根据 (7.3.6) 式可得

$$\hat{\boldsymbol{w}}_{k+1}^{25} = F\hat{\boldsymbol{w}}_k^{25} + Gw_{k+1}$$

这就是 w_t 的预报向量递推公式.

再根据 (7.2.18) 式,并注意 $l > 25$ 时, $\hat{w}_k(l) = 0$, 可得

$$\hat{z}_k(l) = z_{k+l-24} + z_k - z_{k-24} + \sum_{j=1}^{l} \hat{w}_k(j), \quad 1 \leqslant l \leqslant 24$$

$$\hat{z}_k(25) = z_{k-23} + 2(z_k - z_{k-24}) + \sum_{j=1}^{25} \hat{w}_k(j) + \hat{w}_k(1)$$

$$\hat{z}_k(l) = z_{k+l-24r} + r(z_k - z_{k-24}) + \sum_{j=1}^{25}\left(\left[\frac{l-j}{24}\right]+1\right)\hat{w}_k(j), \quad l > 25$$

仍然记 $l^* = l - 24(r-1)$, 则 $1 \leqslant l^* \leqslant 24$, 于是

$$z_{k+l-24r} = z_{k+l^*-24}$$

$$\left[\frac{l-j}{24}\right]+1=\left[\frac{l-24(r-1)+24(r-1)-j}{24}\right]+1$$

$$=\left[\frac{l^*-j}{24}+r-1\right]+1$$

$$=\begin{cases} r, & l^* \geqslant j \\ r-1, & l^* < j \end{cases}$$

因此进一步可表

$$\hat{z}_k(l)=z_{k+l^*-24}+r(z_k-z_{k-24})+r\sum_{j=1}^{l^*}\hat{\omega}_k(j)$$

$$+(r-1)\sum_{j=l^*+1}^{25}\hat{\omega}_k(j)=z_{k+l^*-24}+z_k-z_{k-24}$$

$$+\sum_{j=1}^{l^*}\hat{\omega}_k(j)+(r-1)\left[z_k-z_{k-24}+\sum_{j=1}^{25}\hat{\omega}_k(j)\right]$$

$$=\hat{z}_k(l^*)+(r-1)[\hat{z}_k(25)-\hat{z}_k(1)],\quad l>25$$

这就是 $l>25$ 以后 $\hat{z}_k(l)$ 关于 l 的直接递推公式. 它说明了预报量 $\hat{z}_k(l)$ 随 l 的增加而有明显的周期性和线性发展的趋势: 在相同"相位"上，后一周期的值比前一周期的值多了一项 $[\hat{z}_k(25)-\hat{z}_k(1)]$.

下面给出电力负荷预报的一些数值结果. 我们在第六章§3中曾经用720个电力负荷数据对它的模型进行识别和估计，给出了两种估计模型. 一种是疏系数自回归季节模型(6.3.18)式，它未被检验所接受;另一种是疏系数的滑动平均季节模型(6.3.12)式，它是被检验所接受的. 按照本节例 9 和例 10 介绍的预报方法，我们分别采用上述两种模型对同一个样本序列计算了以下的预报值

$$\begin{matrix} \hat{z}_k^\theta(1), & \hat{z}_k^\theta(25), & \hat{z}_k^\theta(49), \\ \hat{z}_k^\varphi(1), & \hat{z}_k^\varphi(25), & \hat{z}_k^\varphi(49), \end{matrix} \quad k=300,301,\cdots,347$$

其中 $\hat{z}$ 的右上角的记号 θ 或 φ 分别表示按模型(6.3.12)或(6.3.18)式进行预报计算的结果. 为了比较两种不同模型预报的效果，还计算了预报误差量

$$\sigma^{\theta}(l)=\left[\frac{1}{48}\sum_{k=300}^{347}(\hat{z}_k^{\theta}(l)-z_{k+l})^2\right]^{\frac{1}{2}}$$

$$\sigma^{\varphi}(l)=\left[\frac{1}{48}\sum_{k=300}^{347}(\hat{z}_k^{\varphi}(l)-z_{k+l})^2\right]^{\frac{1}{2}} \qquad l=1,\ 25,\ 49$$

它们的具体结果如下

$$\sigma^{\theta}(1)=30.4,\quad \sigma^{\theta}(25)=51.7,\quad \sigma^{\theta}(49)=50.2$$

$$\sigma^{\varphi}(1)=31.6,\quad \sigma^{\varphi}(25)=36.8,\quad \sigma^{\varphi}(49)=33.5$$

从这些数值结果首先可以看出，用模型(6.3.12)做一步预报比(6.3.18)式要好，这是可以预想得到的，因为(6.3.12)式拟合得好．但奇怪的是，在进行较长期的预报时，情况却反过来了．我们认为，这是由于(6.3.18)式关于 z_t 的差分阶数比(6.3.12)式低：前者把 $\nabla_{24}z_t$ 当做平稳序列，而后者把 $\nabla\nabla_{24}z_t$ 当做平稳序列．一般说来，差分阶数越高，较长期的预报效果越差．如上面所表明的，虽然 $\sigma^{\varphi}(1)>\sigma^{\theta}(1)$，但是，$\sigma^{\varphi}(25)$ 和 $\sigma^{\varphi}(49)$ 都远比 $\sigma^{\theta}(25)$，

表 7.3.2　电力负荷的一步预报值

$k+1$	z_{k+1}	$\hat{z}_k^{\varphi}(1)$	$\hat{z}_k^{\theta}(1)$	$k+1$	z_{k+1}	$\hat{z}_k^{\varphi}(1)$	$\hat{z}_k^{\theta}(1)$	$k+1$	z_{k+1}	$\hat{z}_k^{\varphi}(1)$	$\hat{z}_k^{\theta}(1)$
301	855	867.7	854.4	317	747	741.0	738.1	333	824	829.0	841.4
302	871	830.4	856.2	318	842	787.2	790.5	334	758	740.2	768.3
303	868	844.3	864.0	319	837	849.6	844.3	335	675	706.9	717.8
304	869	867.8	863.3	320	875	880.0	886.7	336	680	680.4	698.2
305	858	871.1	891.8	321	867	870.1	864.8	337	606	675.2	670.4
306	871	868.3	882.4	322	818	834.9	846.9	338	599	629.9	630.1
307	886	853.5	870.6	323	735	757.9	763.6	339	603	630.3	617.3
308	870	863.3	871.1	324	748	763.2	769.2	340	650	643.6	620.5
309	867	805.3	832.2	325	725	833.3	811.6	341	697	715.7	686.4
310	755	792.1	788.9	326	830	762.0	775.9	342	783	791.9	745.2
311	711	707.7	727.5	327	811	840.8	800.9	343	837	813.8	788.5
312	700	685.6	721.2	328	817	830.5	803.3	344	918	887.6	855.1
313	676	698.2	692.7	329	876	833.2	831.0	345	887	896.0	867.8
314	672	656.4	672.8	330	889	871.6	857.9	346	896	838.7	853.9
315	680	667.4	672.8	331	890	882.9	867.8	347	825	798.3	801.0
316	687	638.9	686.0	332	882	866.2	870.2	348	807	811.3	829.7

$\sigma^{\theta}(49)$ 小. 此外, $\hat{z}_k^{\varphi}(l)$ 比 $\hat{z}_k^{\theta}(l)$ 容易计算,而在实际使用中,对于较大的 l 计算 $\hat{z}_k(l)$ 还是比较常用到的,所以,虽然(6.3.18)模型未能被检验所接受,我们仍可以采用这个模型进行预报. 最后,在进一步考虑控制问题时,这种模型也有其方便之处(见第九章). 至于前面出现的 $\sigma^{\varphi}(49) < \sigma^{\varphi}(25)$ 和 $\sigma^{\theta}(49) < \sigma^{\theta}(25)$, 这可能是随机误差所致. 一般说来预报步数越大,其误差方差也越大.

我们把 $\hat{z}_k^{\theta}(1)$, $\hat{z}_k^{\varphi}(1)$ 和 z_{k+1} 的数值分别列在表 7.3.2 中供读者参考. 从第六章 §3 的表 6.3.1 可知, $(\hat{\gamma}_0(z))^{\frac{1}{2}} \cong 75.6$, 而 $\sigma^{\theta}(1) = 30.4$, $\sigma^{\varphi}(1) = 31.6$, 这说明上面这两种预报值比用样本平均 $\bar{z}$ 当做预报好得多.

§4. 随机序列的适时预报方法

前面三节讨论了 ARMA 序列的平稳预报,也就是根据全部历史数据 $w_k, w_{k-1}, \cdots$ 给出 w_{k+l} 的线性最小方差估计值,并且将它们推广到 ARIMA 序列和季节性序列的情况. 但是,当实际使用这些原则和方法时,在 $q > 0$ 的情况下,只能计算 $\hat{w}_k(l)$ 的近似值,只有当 k 很大以后,近似误差才可以被忽略. 在某些实际问题中,特别是要求连续进行的适时预报问题,常常希望根据实际上获得的有限个数据 $w_k, w_{k-1}, \cdots, w_1$, 给出 w_{k+l} 的严格的线性最小方差估计,做为它的预报值,并且随 k 的增加给出预报值的递推计算式. 本节将给出一种适应这种需要的新息预报方法.

设 w_t 为 ARMA 序列,并记

$$\mathscr{W}_k^1 = \left\{w:\ w = \sum_{j=1}^{k} \alpha_j w_j,\ \alpha_j \text{ 为实数}\right\}$$

按照第一章 §5 的内容, 根据 $w_k, w_{k-1}, \cdots, w_1$ 对 w_{k+l} 所做的线性最小方差估计,就是 $\mathscr{W}_k^1$ 中的某一元素, 记为

$$\hat{w}_{k+l,k} = E(w_{k+l} | w_k, w_{k-1}, \cdots, w_1) = E(w_{k+l} | \mathscr{W}_k^1)$$

又记一步预报误差为

$$\varepsilon_k = w_k - \hat{w}_{k,k-1} \tag{7.4.1}$$

由于 $\hat{w}_{k,k-1}$ 是从"老"的数据 $w_1, w_2, \cdots, w_{k-1}$ 的线性组合(即

$\mathscr{W}_{k-1}^1$）中所获得的关于 w_k 的最大信息，所以可以说 ε_k 是从新来数据 w_k 中得到的新信息．因此我们称 $\varepsilon_1, \varepsilon_2, \cdots$ 为 $w_1, w_2, \cdots$ 的新息序列．

由 ε_k 和 $\mathscr{W}_k^1$ 的定义可知

$$\mathscr{W}_k^1 = \left\{\varepsilon:\ \varepsilon = \sum_{i=1}^{k} \delta_i\varepsilon_i,\ \delta_i \text{ 为实数}\right\} = \mathscr{E}_k^1 \tag{7.4.2}$$

根据线性最小方差估计的性质（见第一章 §5）可知

$$E\varepsilon_k w_j = 0,\ E\varepsilon_k\varepsilon_j = 0, \qquad j < k \tag{7.4.3}$$

$$\hat{w}_{k+l,k} = \sum_{j=1}^{k} (Ew_{k+l}\varepsilon_j)(E\varepsilon_j^2)^{-1}\varepsilon_j \tag{7.4.4}$$

现在，设 w_t 为 ARMA(p, q)序列，即 $\phi(B)w_t = \theta(B)a_t$，令

$$Q = \max(p, q) \tag{7.4.5}$$

$$\begin{cases} y_t = w_t, & t \leqslant Q \\ y_t = \varphi(B)w_t = w_t - \sum\limits_{j=1}^{p} \varphi_j w_{t-j}, & t > Q \end{cases} \tag{7.4.6}$$

则不难看出

$$\mathscr{Y}_k^1 = \left\{y:\ y = \sum_{j=1}^{k} C_j y_j,\ C_j \text{ 为实数}\right\} = \mathscr{W}_k^1 = \mathscr{E}_k^1 \tag{7.4.7}$$

于是类似于(7.4.4)式便有

$$\hat{y}_{k,k-1} = E(y_k|\mathscr{Y}_{k-1}^1) = \sum_{j=1}^{k-1} J_{kj}\varepsilon_j \tag{7.4.8}$$

其中 $J_{kj} = (Ey_k\varepsilon_j)(E\varepsilon_j^2)^{-1}$ 是实系数．由于当 $k \leqslant Q$ 时 $y_k = w_k$（见(7.4.6)式），所以这时有 $\hat{y}_{k,k-1} = \hat{w}_{k,k-1}$ 及

$$y_k - \hat{y}_{k,k-1} = w_k - \hat{w}_{k,k-1} = \varepsilon_k \tag{7.4.9}$$

而当 $k > Q$ 时，则有

$$\begin{aligned} y_k - \hat{y}_{k,k-1} &= \varphi(B)w_k - E\left(w_k - \sum_{j=1}^{p} \varphi_j w_{k-j} \middle| \mathscr{W}_{k-1}^1\right) \\ &= \varphi(B)w_k - \left(\hat{w}_{k,k-1} - \sum_{j=1}^{k} \varphi_j w_{k-j}\right) \\ &= w_k - \hat{w}_{k,k-1} = \varepsilon_k \end{aligned} \tag{7.4.10}$$

以上两式说明 y_t 和 w_t 有相同的新息序列. 因此，和 (7.4.3)式一样有

$$E\varepsilon_k y_j = 0, \qquad j < k \tag{7.4.11}$$

新息定理[21] 设 w_t 是满足 ARMA(p, q)模型的随机序列，那么

$$\varepsilon_k = \begin{cases} y_1 = w_1, & k = 1 \\ y_k - \sum_{j=1}^{k-1} J_{kj}\varepsilon_j, & 1 < k \leqslant Q \\ y_k - \sum_{j=k-q}^{k-1} J_{kj}\varepsilon_j, & k > Q \end{cases} \tag{7.4.12}$$

其中

$$J_{kj} = \begin{cases} \left[\gamma_{kj}(y) - \sum_{i=1}^{j-1} J_{ji}J_{ki}\gamma_{ii}(\varepsilon)\right][\gamma_{jj}(\varepsilon)]^{-1} \\ \qquad 1 < k \leqslant Q, \ 1 \leqslant j \leqslant k-1 \\ \left[\gamma_{kj}(y) - \sum_{i=k-q}^{j-1} J_{ki}J_{ji}\gamma_{ii}(\varepsilon)\right][\gamma_{jj}(\varepsilon)]^{-1} \\ \qquad k > Q, \ k-q \leqslant j \leqslant k-1 \end{cases} \tag{7.4.13}$$

$$\gamma_{kk}(\varepsilon) = \begin{cases} \gamma_{11}(y), & k = 1 \\ \gamma_{kk}(y) - \sum_{j=1}^{k-1} J_{kj}^2\gamma_{jj}(\varepsilon), & 1 < k \leqslant Q \\ \gamma_{kk}(y) - \sum_{j=k-q}^{k-1} J_{kj}^2\gamma_{jj}(\varepsilon), & k > Q \end{cases} \tag{7.4.14}$$

当求和号的上限小于下限时约定此求和为零值，又记 $\gamma_{kj}(y) = Ey_k y_j$，$\gamma_{jj}(\varepsilon) = E\varepsilon_j^2$.

证明　由 (7.4.9)，(7.4.10) 和 (7.4.11) 式可知

$$E(y_k - \hat{y}_{k,k-1})y_j = E\varepsilon_k y_j = 0, \quad 1 \leqslant j < k$$

当 $t > Q$ 时 y_t 为 $q(q \leqslant Q)$ 阶滑动平均，且 $y_t = \theta(B)a_t$，所以当 $k > Q$ 时，

$$Ey_k y_j = 0, \quad 1 \leqslant j < k - q$$

这说明 y_k 与 $\mathscr{Y}_{k-q-1}^1 = \mathscr{E}_{k-q-1}^1$ 不相关，所以当 $k > Q$ 时

$$E y_k \varepsilon_j = 0, \quad 1 \leqslant j < k - q$$

以此代入(7.4.8)式可知

$$\hat{y}_{k,k-1} = \sum_{j=k-q}^{k-1} J_{kj} \varepsilon_j, \qquad k > Q$$

由此式和(7.4.8)式即知(7.4.12)式成立. 再应用(7.4.3)式，从(7.4.12)式立刻可得(7.4.14)式.

下面证明(7.4.13)式. 先求 $E\varepsilon_i y_j$. 当 $i \leqslant Q$ 时

$$E\varepsilon_i y_j = E\varepsilon_i\left(\varepsilon_j + \sum_{s=1}^{j-1} J_{js}\varepsilon_s\right) = \begin{cases} J_{ji}\gamma_{ii}(\varepsilon), & 1 \leqslant i \leqslant j-1 \\ \gamma_{jj}(\varepsilon), & i = j \\ 0, & i > j \end{cases}$$

当 $i > Q$ 时

$$E\varepsilon_i y_j = E\varepsilon_i\left(\varepsilon_j + \sum_{s=j-q}^{j-1} J_{js}\varepsilon_s\right)$$

$$= \begin{cases} J_{ji}\gamma_{ii}(\varepsilon), & j-q \leqslant i \leqslant j-1 \\ \gamma_{jj}(\varepsilon), & i = j \\ 0, & i > j \text{ 或 } 1 \leqslant i < j-q \end{cases}$$

根据以上两式和(7.4.12)式可知，当 $1 \leqslant j < k \leqslant Q$ 时

$$J_{kj}\gamma_{jj}(\varepsilon) = E\varepsilon_j y_k = E\left(y_j - \sum_{i=1}^{j-1} J_{ji}\varepsilon_i\right) y_k$$

$$= \gamma_{kj}(y) - \sum_{i=1}^{j-1} J_{ji} J_{ki} \gamma_{ii}(\varepsilon)$$

当 $k > Q$ 且 $k - q \leqslant j < k - 1$ 时，

$$J_{kj}\gamma_{jj}(\varepsilon) = E\varepsilon_j y_k = E\left(y_j - \sum_{i=\max(j-q,1)}^{j-1} J_{ji}\varepsilon_i\right) y_k$$

$$= \gamma_{kj}(y) - \sum_{i=\max(j-q,1)}^{j-1} J_{ji} J_{ki} \gamma_{ii}(\varepsilon)$$

将以上两式两端乘以 $[\gamma_{jj}(\varepsilon)]^{-1}$ 便得到(7.4.13)式. 定理证明完毕.

用(7.4.12)，(7.4.13)和(7.4.14)式即可由 y_k(或 w_k)递推地求出新息序列 ε_k，为详尽起见，我们作几点补充说明：

1. $\gamma_{kj}(y)$可以直接计算. 当 $k \leqslant Q$ 时，$y_k = w_k$，即 $\varphi(B)y_k = \theta(B)a_k$；当 $k > Q$ 时，$y_k = \varphi(B)w_k = \theta(B)a_k$ 是 q 阶滑动平均. 根据 y_k 与 w_k 的这种关系，可以利用 $\gamma_j(\gamma_j = Ew_tw_{t-j})$ 和模型参数 $\theta_1, \theta_2, \cdots, \theta_q$ 等值求出 $\gamma_{kj}(y)$ 来. 还须注意，当 k，$j > Q$ 以后，$\gamma_{kj}(y)$ 只是 $k-j$ 的函数.

例如，求 $(1-\varphi_1 B)w_t = (1-\theta_1 B)a_t$ 中的 $\gamma_{kj}(y)$，此时 $p = q = Q = 1$，由第二章 §2 例 6 可知

$$\gamma_0 = \frac{1+\theta_1^2 - 2\theta_1\varphi_1}{1-\varphi_1^2}\sigma_a^2, \quad \gamma_1 = \frac{(1-\theta_1\varphi_1)(\varphi_1-\theta_1)}{1-\varphi_1^2}\sigma_a^2$$

于是

$$\gamma_{11}(y) = Ey_1^2 = Ew_1^2 = \gamma_0$$
$$\gamma_{22}(y) = Ey_2^2 = E(a_2 - \theta_1 a_1)^2 = (1+\theta_1^2)\sigma_a^2$$
$$\gamma_{21}(y) = Ey_2y_1 = E(a_2 - \theta_1 a_1)w_1 = \gamma_1 - \varphi_1\gamma_0 = -\theta_1\sigma_a^2$$
$$\gamma_{kk}(y) = Ey_k^2 = E(a_k - \theta_1 a_{k-1})^2 = \gamma_{22}(y)$$
$$\gamma_{kk-1}(y) = Ey_ky_{k-1} = E(a_k - \theta_1 a_{k-1})(a_{k-1} - \theta_1 a_{k-2}) = -\theta_1\sigma_a^2$$
$$\gamma_{kk-j}(y) = Ey_ky_{k-j} = E(a_k - \theta_1 a_{k-1})y_{k-j} = 0, \quad j > 1, \ k \geqslant 2$$

2. 利用(7.4.13)，(7.4.14)式和 $\gamma_{kj}(y)$，便可递推计算出 J_{kj} 和 $\gamma_{kk}(\varepsilon)$，即

$$\gamma_{11}(\varepsilon) = \gamma_{11}(y)$$
$$J_{21} = \gamma_{21}(y)[\gamma_{11}(\varepsilon)]^{-1}, \quad \gamma_{22}(\varepsilon) = \gamma_{22}(y) - J_{21}^2\gamma_{11}(\varepsilon)$$
$$J_{31} = \gamma_{31}(y)[\gamma_{11}(\varepsilon)]^{-1}$$
$$J_{32} = [\gamma_{32}(y) - J_{31}J_{21}\gamma_{11}(\varepsilon)][\gamma_{22}(\varepsilon)]^{-1}$$
$$\gamma_{33}(\varepsilon) = \gamma_{33}(y) - [J_{31}^2\gamma_{11}(\varepsilon) + J_{32}^2\gamma_{22}(\varepsilon)]$$
$$\cdots\cdots\cdots$$

当 $k < Q$ 时，须求出 $J_{k1}, J_{k2}, \cdots, J_{kk-1}$；当 $k > Q$ 以后，只要求出 $J_{kk-q}, J_{kk-q+1}, \cdots, J_{kk-1}$ 即可. 由此可见用递推公式求 J_{kj} $(k-q \leqslant j \leqslant k-1)$ 和 $\gamma_{kk}(\varepsilon)$ 的最大计算量只和 q 的大小有关，而和 k 无关，这就是说它不会随 k 的增长而增加.

3. 当计算出 $J_{kj}\,(k-q\leqslant j\leqslant k-1)$ 以后，再用(7.4.6)和(7.4.12)式便可求出 ε_k 了，即

$$
\begin{aligned}
\varepsilon_1 &= y_1 \\
\varepsilon_2 &= y_2 - J_{21}\varepsilon_1, \\
\varepsilon_3 &= y_3 - J_{31}\varepsilon_1 - J_{32}\varepsilon_2 \\
&\cdots\cdots
\end{aligned}
$$

若要仔细讨论，仍须分 $k\leqslant Q$ 和 $k>Q$ 两种情况，具体叙述与上边的内容相似．

现在讨论如何利用已获得的新息序列递推公式，给出适时预报的公式．根据(7.4.6)式和(7.4.12)式，

$$
\varepsilon_{k+l}=\begin{cases} w_{k+l}-\sum\limits_{j=1}^{k+l-1} J_{k+l,j}\varepsilon_j, & k+l\leqslant Q \\ w_{k+l}-\sum\limits_{j=1}^{p}\varphi_j w_{k+l-j}-\sum\limits_{j=k+l-q}^{k+l-1} J_{k+l,j}\varepsilon_j, & k+l>Q \end{cases}
$$

将上式右边的求和项都移到左边来，然后求这些量在 $\mathscr{W}_k^1$ 已知下的线性最小方差估计．由此得到

$$
\hat{w}_{k+l,k}=\begin{cases} \sum\limits_{j=1}^{k} J_{k+l,j}\varepsilon_j, & k+l\leqslant Q \\ \sum\limits_{j=1}^{p}\varphi_j\hat{w}_{k+l-j,k}+\sum\limits_{j=k+l-q}^{k} J_{k+l,j}\varepsilon_j, & k+l>Q \end{cases} \tag{7.4.15}
$$

其中当 $l-j\leqslant 0$ 时，$\hat{w}_{k+l-j,k}=w_{k+l-j}$；当求和号的上限小于下限时令其求和为零．要注意，$J_{kj}$ 只与 w_t 的模型参数值有关，而与它的样本值无关，ε_k 还与 w_t 的样本值 $w_1, w_2, \cdots, w_k$ 有关．

当 $k+l>Q$ 时，由(7.4.15)式可得

$$
\varphi(B)\hat{w}_{k+l,k}=0, \qquad l>q \tag{7.4.16}
$$

其中后移算子 B 作用在 l 上．特别在 AR$(p,0)$ 序列的情况下，由于这时 $q=0$，$Q=\max(p,q)=p$，于是(7.4.16)式为，

$$
\varphi(B)\hat{w}_{k+l,k}=0, \qquad l>0, \quad k\geqslant p
$$

而且初始值 $\hat{w}_{k+j,k}=w_{k+j}\ (j=0,-1,\cdots,-p+1)$，这表明

此时 $\hat{w}_{k+l,k}$ 与 $\hat{w}_k(l)$ 完全一致，即当 $k>p$ 以后，AR 序列的适时预报和平稳预报是完全相同的. 因此 AR$(p,0)$ 序列的新息序列为

$$\varepsilon_k=w_k-\hat{w}_{k,k-1}=w_k-\hat{w}_{k-1}(1)=a_k,\quad k>p \quad (7.4.17)$$

在 w_t 为 MA 或 ARMA 序列的时候，一般 $\varepsilon_k\neq a_k$；但是，由于 $\hat{w}_{k,k-1}$，$\hat{w}_{k-1}(1)$ 分别是在已知 $\mathscr{W}^1_{k-1}$ 和 $\mathscr{W}_{k-1}$ 时的线性最小方差估计，显然 $\mathscr{W}^1_{k-1}\subset\mathscr{W}_{k-1}$，所以

$$\begin{aligned}\sigma_a^2&=E[w_k-\hat{w}_{k-1}(1)]^2=E\left[w_k-\sum_{j=1}^{\infty}\pi_j w_{k-j}\right]^2\\&\leqslant E(w_k-\hat{w}_{k,k-1})^2=E\varepsilon_k^2\leqslant E\left(w_k-\sum_{j=1}^{k-1}\pi_j w_{k-j}\right)^2\\&=E\left(a_k+\sum_{j=k}^{\infty}\pi_j w_{k-j}\right)^2=\sigma_a^2+E\left(\sum_{j=k}^{\infty}\pi_j w_{k-j}\right)^2\end{aligned}$$

由于 $|\pi_j|\leqslant C_1e^{-C_2|j|}$，所以上式右边最后一个求和项的方差将随 k 增长而趋于零，由此便得到

$$\sigma_a^2=E[w_k-\hat{w}_k(1)]^2=\lim_{k\to\infty}E\varepsilon_k^2$$

另外，由 (7.4.1) 式又知

$$E(a_k\varepsilon_k)=Ea_k(w_k-\hat{w}_{k,k-1})=Ea_kw_k=\sigma_a^2$$

于是由以上二式可知，当 $k\to\infty$ 时

$$E(a_k-\varepsilon_k)^2=Ea_k^2+E\varepsilon_k^2-2Ea_k\varepsilon_k=E\varepsilon_k^2-\sigma_a^2\to 0$$

由此可知，当 $k\to\infty$时，不仅平稳预报与适时预报的误差方差渐近相等，而且误差本身也是渐近相等的. 对于 $l(l\geqslant 1)$ 步预报也有类似性质，我们不一一证明了.

上述分析和证明告诉我们，无论对 AR，MA 或 ARMA 序列，当 k 很大以后，适时预报与平稳预报都渐近趋于一致. 因此，在实际应用中遇到连续预报的问题时，若要求从较少的数据开始预报，并且希望尽可能给出好的预报值，那末开始一段，可以利用本节的适时预报方法，而当 k 增长到一定数量以后，为了减少每步的计算量，再改用平稳预报的递推方法. 这样把两种方法结合起来使用，

既可以提高预报精度，又可以节省计算量.

关于 ARIMA 序列及季节性序列的适时预报方法，仍然是从它们与平稳序列的关系出发，利用平稳序列的适时预报组成这些序列的预报公式.

例如，$\nabla z_t = w_t$，w_t 为 ARMA 序列. 那么，定义 $\hat{z}_{k+l,k} = E(z_{k+l}|z_1, z_2, \cdots, z_k) = E(z_{k+l}|z_1, w_2, \cdots, w_k)$，由此，我们可以导出类似于 (7.2.13) 式的公式

$$\begin{aligned}\hat{z}_{k+l,k} &= E\left(z_k + \sum_{j=1}^{l} w_{k+j}|z_1, w_2, \cdots, w_k\right) \\ &= z_k + E\left(\sum_{j=1}^{l} w_{k+j}|z_1, w_2, \cdots, w_k\right)\end{aligned}$$

我们仍假定初值 z_1 与 w_t 独立，于是

$$\begin{aligned}\hat{z}_{k+l,k} &= z_k + E\left(\sum_{j=1}^{l} w_{k+j}|w_2, \cdots, w_k\right) \\ &= z_k + \sum_{j=1}^{l} \hat{w}_{k+j,k}, \qquad k \geqslant 2 \end{aligned} \tag{7.4.18}$$

其中 $\hat{w}_{k+j,k} = E(w_{k+j}|w_2, w_3, \cdots, w_k) \equiv E(w_{k+j}|\mathscr{W}_k^2)$，它可由前面的适时预报公式递推计算.

又如，$\nabla_s z_t = w_t$，w_t 为 ARMA 序列. 那么由类似上面的分析可以得到与 (7.2.15) 式相似的公式:

$$\hat{z}_{k+l,k} = z_{k+l-rs} + \sum_{j=0}^{r-1} \hat{w}_{k+l-js,k}, \qquad k > s+1 \tag{7.4.19}$$

其中 $(r-1)s < l \leqslant rs$，$\hat{w}_{k+l-js,k} = E(w_{k+l-js}|\mathscr{W}_k^{s+1}) = E(w_{k+l-js}|w_{s+1}, w_{s+2}, \cdots, w_k)$ 它可由前面的适时预报公式递推计算.

不过实际中的季节性模型的预报问题，采样间隔都是比较长的，所以一般说来并不使用适时预报方法.

第八章　线性系统的识别及参数估计

§1. 离散线性系统

在很多控制系统中，常常以电器元件组成的线路做为输入输出系统的例子．如图 8.1.1 所示：



图 8.1.1　输入输出系统

输入 $x(t)$ 和输出 $y(t)$ 都是随时间连续变化的电讯号(如电压量)，系统结构是由确定的电子线路组成的．例如在绪论中提到的轧钢过程就是一例．但是，对于大量的实际系统来说，不能完全知道其结构，或者无法知道，对输入输出量却能进行测量．这时，为了掌握 $y(t)$ 和 $x(t)$ 的依赖关系，提出了如何依据输入输出量的测量数据，对系统进行描述和识别的问题．近年来由于电子计算技术的广泛使用，用连续系统离散化的方法解决系统描述和识别等问题更被人们注意了．离散化的办法常用等间隔瞬时采样，等间隔积分采样等．当然，还有的实际系统本身就是离散系统，例如第二章 §1 中例 1 的计算机的舍入误差系统．这一章将讨论用 ARMA 序列分析的技巧，解决离散系统的描述与识别问题．为此在本节先介绍一些关于离散系统的预备知识．

下面所称的系统，限于考虑标量离散实变系统，即它的输入输出都是数值随机序列．对于某个系统，当以任意两个不同的 $x_t^{(1)}$ 及 $x_t^{(2)}$ 为输入时，我们把其相应的输出记为 $y_t^{(1)}$ 或 $y_t^{(2)}$；如果以 $x_t^{(1)}$ 和 $x_t^{(2)}$ 的任意线性组合 $\alpha x_t^{(1)}+\beta x_t^{(2)}$ 为系统的输入时，其输出必是 $\alpha y_t^{(1)}+\beta y_t^{(2)}$，那么称此系统为线性系统．又若对任意 τ 值，以

$x_{t-\tau}$ 为输入时，其输出为 $y_{t-\tau}$，则称此系统为非时变或定常系统。以下所讨论的系统，都是这种非时变的线性系统。

如果一个系统的输入输出关系为

$$y_t = \sum_{j=-\infty}^{\infty} h_j x_{t-j} \tag{8.1.1}$$

其中 $\sum |h_j| < \infty$，那么它必定是线性非时变系统。如果进一步有

$$y_t = \sum_{j=0}^{\infty} h_j x_{t-j} \tag{8.1.2}$$

即 $h_j = 0\ (j < 0)$，则称此系统为物理可实现的．这时，它的输出量只依赖于现在和过去时刻的输入量．

在 (8.1.2) 式中，当输入为单位“脉冲”，即 $x_t = \delta_t$ $(\delta_0 = 1;\ \delta_t = 0,\ t \neq 0)$ 时，系统相应的输出为 $y_t = h_t$．类似于连续系统，我们称 h_t 为系统 (8.1.2) 的脉冲响应函数．又称

$$h_t^* = \sum_{j=0}^{t} h_j \tag{8.1.3}$$

为阶跃响应函数．

如果令

$$h(B) = \sum_{j=0}^{\infty} h_j B^j \tag{8.1.4}$$

则 (8.1.2) 式又可写做

$$y_t = h(B) x_t \tag{8.1.5}$$

通常称 $h(B)$ 为传递函数，(8.1.5)式为传递形式．

我们将要用到的离散系统是 (8.1.2) 式的一种重要的特殊类型，即输入输出过程满足如下的差分方程式：

$$(1 + \xi_1 \nabla + \xi_2 \nabla^2 + \cdots + \xi_r \nabla^r) y_t = c(1 + \eta_1 \nabla + \cdots + \eta_s \nabla^s) x_{t-b} \tag{8.1.6}$$

其中 $\xi_i (1 \leqslant i \leqslant r)$，$\eta_j (1 \leqslant j \leqslant s)$ 是常数系数，∇^i 是 i 阶差分运算．因为 $\nabla = I - B$，上式又容易改写成

$$(1-\zeta_1 B-\cdots-\zeta_r B^r)y_t=(\omega_0-\omega_1 B-\cdots-\omega_s B^s)x_{t-b} \quad (8.1.7)$$

这样的系统称为有限阶离散定常线性系统．它的阶数用 (r, s, b) 表示；ζ_i, ω_j 是它的参数；$b(\geqslant 0)$ 是它的延迟步数．令

$$\zeta(B)=1-\zeta_1 B-\cdots-\zeta_r B^r$$

$$\Omega(B)=\omega_0-\omega_1 B-\cdots-\omega_s B^s$$

$$\omega(B)=\Omega(B)B^b$$

于是 (8.1.7) 式又可简记为

$$\zeta(B)y_t=\omega(B)x_t \quad (8.1.8)$$

为了求出此系统的传递函数和脉冲响应函数，将 (8.1.5) 式代入上式得

$$\zeta(B)h(B)x_t=\omega(B)x_t$$

对于任何输入量 x_t，上式均成立．于是传递函数 $h(B)$ 满足方程：$\zeta(B)h(B)=\omega(B)$．对比两边系数，易知脉冲响应函数 h_t 满足下式：

$$h_t=\begin{cases}0, & t<b\\ \omega_0, & t=b\\ \zeta_1 h_{t-1}+\cdots+\zeta_r h_{t-r}-\omega_{t-b}, & b+1\leqslant t\leqslant b+s\\ \zeta_1 h_{t-1}+\cdots+\zeta_r h_{t-r}, & t\geqslant b+s+1\end{cases} \quad (8.1.9)$$

为了保证由上式解出的 h_t 具有绝对收敛性，即 $\sum|h_j|<\infty$，我们只讨论 $\zeta(\lambda)=0$ 的根都在单位圆外的情况．这时此系统被称为稳定系统．这个条件相当于 ARMA 序列的平稳性条件，系统参数的稳定域也和 ARMA 模型参数的平稳域类似，可参看第二章 §2 的讨论．在稳定性条件下，(8.1.8) 式的系统输出可表为

$$y_t=\omega(B)\zeta^{-1}(B)x_t=h(B)x_t=\sum_{j=0}^{\infty}h_j x_{t-j} \quad (8.1.10)$$

其中 h_t 由差分方程 (8.1.9) 式算出．现在分析一下这种 h_t 序列的变化规律，在 b 以前，$h_t=0$；从 b 到 $b+s$，h_t 呈现不规则性；从 $b+s$ 以后，h_t 满足齐次差分方程

$$h_t-\zeta_1 h_{t-1}-\cdots-\zeta_r h_{t-r}=0, \quad t>b+s \quad (8.1.11)$$

由于 $\zeta(B)$ 满足稳定性条件，所以必有

$$|h_t| \leqslant c_1 e^{-c_2 t}$$

c_1, c_2 是适当的正常数，即 h_t 被负指数函数所控制. 特别当 $r=0$ 时，h_t 从 $b+s+1$ 开始截尾；当 $r=1$ 时，$|h_t|$ 从 $b+s+1$ 开始为负指数型的函数；而且，若 $\zeta_1>0$，则 h_t 从此恒为正或负，若 $\zeta_1<0$，则 h_t 从此正负交错地取值；当 $r=2$ 时，h_t 的尾部 $(t \geqslant b+s+1)$ 与第二章 §2 例 1 的 ρ_k 的尾部完全相似.

对于 (8.1.7) 式这样的系统，如果取得了一段输入输出的数据序列，如何用它们对系统进行识别估计，是这一章所要讨论的主要内容. 常见的识别方法有两种：一种是以脉冲量做为输入，那么相应的输出便是脉冲响应函数 h_t；另一种是以白噪声序列做为输入，然后对输出量进行时序分析，由此得出的估计模型便是对 (8.1.7) 式的估计. 使用这两种方法时，要求实际系统能够以脉冲或白噪声为输入进行试验. 但是，在某些实际问题中，这是做不到的，或者这样做的代价太高了. 这里我们将要介绍一种新的方法，它对输入序列不必加严格的限制，即在系统正常运行时，只要记录下输入量和输出量就可以估计系统的阶数和参数. 具体地说，即当输入 x_t 可以设想为平稳序列的场合，这时我们可认为输入 x_t 为某一 ARMA 序列，或者能用 ARMA 模型来拟合，即

$$\varphi(B)x_t = \theta(B)a_t \tag{8.1.12}$$

其中 a_t 为如前几章所述的白噪声. 那么，将它代入 (8.1.7) 式便得到

$$\begin{aligned}\zeta(B)\varphi(B)y_t = \omega(B)\varphi(B)x_t = \omega(B)\theta(B)a_t \\ = \Omega(B)\theta(B)a_{t-b}\end{aligned} \tag{8.1.13}$$

我们总假定 $\zeta(B)\varphi(B)$ 与 $\Omega(B)\theta(B)$ 无公共因子，而且 $\omega_0 \neq 0$；由此可见，输出 y_t 为 ARMA $(p+r, q+s)$ 序列，与它相应的白噪声序列为 $\omega_0 a_{t-b}$，它相对于 a_t 而言有 b 步延迟，并且相差一个因子 ω_0.

当实际输入 x_t 为非平稳序列时，我们假定它能用某一 ARIMA (p, d, q) 序列描述，即

$$\varphi(B)\nabla^d x_t = \theta(B)a_t \tag{8.1.14}$$

这时由(8.1.7)式得到

$$\zeta(B)\varphi(B)\nabla^d y_t = \omega(B)\theta(B)a_t = \Omega(B)\theta(B)a_{t-b} \tag{8.1.15}$$

同样假定 $\zeta(B)\varphi(B)\nabla^d$ 与 $\Omega(B)\theta(B)$ 无公共因子，而且 $\omega_0 \neq 0$；可见 y_t 为 ARIMA $(p+r, d, q+s)$ 序列，相应的白噪声序列为 $\omega_0 a_{t-b}$.

我们还可以进一步考虑如下的更有实际意义的情形，即假设在记录 y_t 时，带有某种干扰量 n_t(比如记录误差，或模型近似产生的各种误差). 这时，真正取得的输出量为

$$z_t = y_t + n_t \tag{8.1.16}$$

其中设 n_t 为 ARMA (p_n, q_n) 序列，而且与 x_t 相互独立. 当然也可以考虑 n_t 为 ARIMA 序列，但是并没有本质差别. 我们只讨论 n_t 为平稳序列的情况，即

$$\varphi_n(B)n_t = \theta_n(B)\alpha_t$$

α_t 为白噪声. 于是(8.1.7)式在输出有干扰时的模型被描述为

$$z_t = y_t + n_t = \zeta^{-1}(B)\omega(B)x_t + \varphi_n^{-1}(B)\theta_n(B)\alpha_t \tag{8.1.17}$$

若 x_t 为 ARMA (p, q) 序列，则由上式得到

$$z_t = \zeta^{-1}(B)\omega(B)\varphi^{-1}(B)\theta(B)a_t + \varphi_n^{-1}(B)\theta_n(B)\alpha_t \tag{8.1.18}$$

它可看成是以 a_t 和 α_t 为双输入的输出，如图 8.1.2 所示. (8.1.18)式的右方第一项称为系统的传递部分，第二项称为系统的噪声部分. 若 x_t 为 ARIMA (p, d, q) 序列，则(8.1.18)式可写为

$$\nabla^d z_t = \zeta^{-1}(B)\omega(B)\varphi^{-1}(B)\theta(B)a_t + \varphi_n^{-1}(B)\theta_n(B)\nabla^d\alpha_t \tag{8.1.19}$$

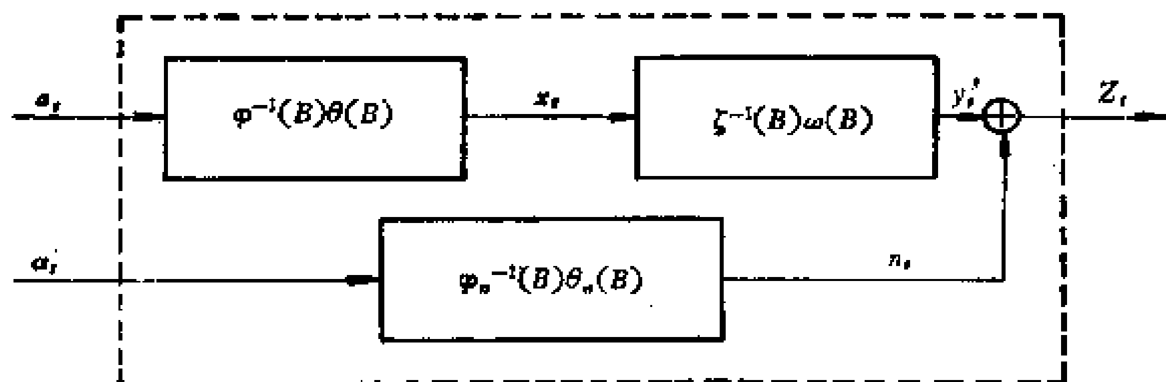


图8.1.2 带干扰的有限阶系统

经过这样差分处理后，这类非平稳输入的系统识别问题也可以化为平稳输入的情况来解决.

如上所述，我们所说的系统(带干扰的线性定常稳定有限阶系统)识别和估计的全部内容可归结为，利用输入输出数据列$(x_1, z_1), (x_2, z_2), \cdots, (x_N, z_N)$,对系统阶数 (r, s, b), (p_n, q_n)和 (p, d, q) 进行识别,以及对诸参数 $\zeta_i(1 \leqslant i \leqslant r)$, $\omega_j(0 \leqslant j \leqslant s)$, $\varphi_{ni}(1 \leqslant i \leqslant p_n)$,$\theta_{nj}(1 \leqslant j \leqslant q_n)$, $\varphi_i(1 \leqslant i \leqslant p)$ 和 $\theta_j(1 \leqslant j \leqslant q)$ 进行估计.

§2. 样本互相关函数

从前一节可以看到,为了进行系统识别,需要处理两个随机序列的样本数据. 如果把一个系统的输入 x_t 和输出 z_t 看做是均值为零的二维平稳随机向量序列，那么，按照第一章§4所介绍的内容,这个二维随机序列的协方差函数应是一个二阶矩阵,它的右上角元素是

$$\gamma_k(xz) = Ex_tz_{t+k}, \qquad k = 0, \pm 1, \pm 2, \cdots \tag{8.2.1}$$

称它为 x_t 与 z_t 的互协方差函数. 另外,由平稳性容易看出

$$\gamma_k(zx) = Ez_tx_{t+k} = Ex_{t+k}z_t = Ex_tz_{t-k} = \gamma_{-k}(xz) \tag{8.2.2}$$

为简明起见,令 $\sigma_x^2 = \gamma_0(xx)$, $\sigma_z^2 = \gamma_0(zz)$, 并令

$$\rho_k(xz) = \frac{\gamma_k(xz)}{\sigma_x\sigma_z}, \qquad \rho_k(zx) = \frac{\gamma_k(zx)}{\sigma_x\sigma_z} \tag{8.2.3}$$

称它为 x_t 与 z_t (z_t 与 x_t) 的互相关函数. 同样易知

$$\rho_k(xz) = \rho_{-k}(zx) \tag{8.2.4}$$

一般地,设 $(x_1, z_1), (x_2, z_2), \cdots, (x_N, z_N)$ 是任一均值为零的二维平稳随机序列的一段样本值,N 为样本长度. 令

$$\hat{\gamma}_k(xz) = \begin{cases} \dfrac{1}{N}\sum\limits_{j=1}^{N-k} x_jz_{j+k}, & k \geqslant 0 \\ \dfrac{1}{N}\sum\limits_{j=1}^{N+k} x_{j-k}z_j, & k < 0 \end{cases} \tag{8.2.5}$$

类似地可以定义 $\hat{\gamma}_k(zx)$. 我们称它们为 x_t 与 z_t (或 z_t 与 x_t)的样

本互协方差函数．显然，若在上式中以 x 代替 z（或以 z 代替 x），就得到 x_t（或 z_t）的样本自协方差函数．为了简化符号我们记

$$\hat{\gamma}_k(xx)=\hat{\gamma}_k(x),\quad \hat{\gamma}_k(zz)=\hat{\gamma}_k(z),\quad \hat{\sigma}_x^2=\hat{\gamma}_0(x),\quad \hat{\sigma}_z^2=\hat{\gamma}_0(z)$$

又令

$$\hat{\rho}_k(xz)=\frac{\hat{\gamma}_k(xz)}{\hat{\sigma}_x\hat{\sigma}_z},\qquad \hat{\rho}_k(zx)=\frac{\hat{\gamma}_k(zx)}{\hat{\sigma}_x\hat{\sigma}_z}\tag{8.2.6}$$

称为 x_t 与 z_t（z_t 与 x_t）的样本互相关函数．易见，

$$\hat{\gamma}_k(xz)=\hat{\gamma}_{-k}(zx),\qquad \hat{\rho}_k(xz)=\hat{\rho}_{-k}(zx)\tag{8.2.7}$$

完全类似地有 $\hat{\rho}_k(xx)=\hat{\rho}_k(x)$，$\hat{\rho}_k(zz)=\hat{\rho}_k(z)$；它们分别是 x_t 和 z_t 的样本自相关函数．

我们引入样本互相关函数的目的，是为了用它来估计理论互相关函数，因此很自然地要关心它与理论值的近似程度如何．可以证明，在本书所限定的正态平稳序列的条件下（事实上还可以在更一般条件下），这些估计是渐近无偏和相容的．证明方法和一维情形相类似．这些估计量的渐近方差由下述多维情形的 Bartlett 公式给出，即当样本 N 较大时，

$$\begin{aligned}\operatorname{cov}(\hat{\rho}_k(xz),\hat{\rho}_{k+l}(xz))\approx\frac{1}{N}\sum_{j=-\infty}^{\infty}\Big\{&\rho_j(x)\rho_{j+l}(z)\\&+\rho_{-j}(xz)\rho_{j+2k+l}(xz)+\rho_k(xz)\rho_{k+l}(xz)\\&\left[(\rho_j(xz))^2+\frac{1}{2}(\rho_j(x))^2+\frac{1}{2}(\rho_j(z))^2\right]\\&-\rho_k(xz)[\rho_j(x)\rho_{j+k+l}(xz)+\rho_{-j}(xz)\rho_{j+k+l}(z)]\\&-\rho_{k+l}(xz)[\rho_j(x)\rho_{j+k}(xz)+\rho_{-j}(xz)\rho_{j+k}(z)]\Big\}\end{aligned}\tag{8.2.8}$$

当 $l=0$ 时，由上式即得到 $\hat{\rho}_k(xz)$ 的误差方差 $\operatorname{var}(\hat{\rho}_k(xz))$ 的渐近公式；当 $x_t=z_t$ 时，$\rho_j(x)=\rho_j(z)=\rho_j(xz)$，以此代入上式便得到一维的 Bartlett 公式（证明见附录 §2）．

如果存在两个整数（包括 $\pm\infty$）$K_1\leqslant K_2$，使得当 $j>K_2$ 或 $j<K_1$ 时，恒有 $\rho_j(xz)=0$，那么当 k，$k+l$ 和 $k+\frac{l}{2}$ 都在区间 $[K_1, K_2]$ 之外时，由（8.2.8）式可得

$$\mathrm{cov}(\hat{\rho}_k(xz),\ \hat{\rho}_{k+l}(xz)) \approx \frac{1}{N}\sum_{j=-\infty}^{\infty}\rho_j(x)\rho_{j+l}(z) \qquad (8.2.9)$$

特别当 k 在 $[K_1, K_2]$ 之外且 $l=0$ 时，

$$\mathrm{var}(\hat{\rho}_k(xz)) \approx \frac{1}{N}\sum_{j=-\infty}^{\infty}\rho_j(x)\rho_j(z) \qquad (8.2.10)$$

如果除上述条件外，再设 x_t 是白噪声，则以上两式进一步简化成

$$\mathrm{cov}(\hat{\rho}_k(xz),\ \hat{\rho}_{k+l}(xz)) \approx \frac{1}{N}\rho_l(z) \qquad (8.2.11)$$

$$\mathrm{var}(\hat{\rho}_k(xz)) \approx \frac{1}{N} \qquad (8.2.12)$$

现在，设 x_t 是白噪声，且对任意 j，$\rho_j(xz)=0$，应用上述结果可知，此时 $\hat{\rho}_k(xz)$（当 k 变化时）做为随机序列，其自相关函数近似为

$$\rho(\hat{\rho}_k(xz),\ \hat{\rho}_{k+l}(xz)) = \frac{\mathrm{cov}(\hat{\rho}_k(xz),\ \hat{\rho}_{k+l}(xz))}{[\mathrm{var}(\hat{\rho}_k(xz))\cdot\mathrm{var}(\hat{\rho}_{k+l}(xz))]^{\frac{1}{2}}} \approx \rho_l(z) \qquad (8.2.13)$$

由此可见在所设条件下，x_t 与 z_t 的样本互相关函数的方差近似为 $\frac{1}{N}$，而且样本互相关函数的理论自相关函数渐近等于 z_t 的自相关函数．这些性质在本章以下几节中将要用到．

§3. 系统的初步识别与估计

我们先从有输出干扰噪声的模型（8.2.17）式讨论起．如前所述，假定 n_t 为 ARMA (p_n, q_n) 序列，x_t 为 ARIMA (p, d, q) 序列．当 $d>0$ 时，可求出 $\nabla^d x_t$ 以代替 z_t，于是 $\nabla^d z_t$ 的模型也变成（8.1.17）式的形式（如（8.1.19）式）．因此，我们只须讨论 x_t 为 ARMA (p, q) 序列的情况．为了解决本章 § 1 末尾所指出的识别和估计问题，可分为两步进行：第一步是对 x_t 的模型进行识别和估计，这只须用到数据 $x_1, x_2, \cdots, x_N$ 和以前几章的方法，这里不再重复．第二步，即在 x_t 的模型和参数被确定以后，进一步识

别估计(8.1.18)式中的$\zeta(B)$，$\omega(B)$和$\varphi_n(B)$，$\theta_n(B)$的阶数与参数．为解决这后一步的问题，我们先将(8.1.18)式改写成

$$\varphi(B)\theta^{-1}(B)z_t = \zeta^{-1}(B)\omega(B)a_t + \varphi(B)\theta^{-1}(B)n_t \qquad (8.3.1)$$

令 $u_t = \varphi(B)\theta^{-1}(B)z_t$，$\varepsilon_t = \varphi(B)\theta^{-1}(B)n_t$，于是(8.3.1)式可简记为

$$u_t = \zeta^{-1}(B)\omega(B)a_t + \varepsilon_t = h(B)a_t + \varepsilon_t \qquad (8.3.2)$$

这样一来，对模型(8.3.1)式的识别和估计问题，便等同于对模型(8.3.2)式的识别和估计问题．后一模型形式上和原来的(8.1.17)式相似，不过它的输入换成了白噪声序列a_t．这种转化手段称为"预白化"方法．当然，如本章§1所说过的那样，若能对原系统直接输入白噪声a_t，这一转化步骤便可省去．由此可见，在输入为白噪声时，系统(8.1.17)式的识别与估计可以直接使用本章以下将要介绍的方法．

如前所述，我们已假定x_t与n_t相互独立，因而，a_t与a_t，ε_t也是相互独立的序列．于是，在(8.3.2)式两端乘以a_{t-k}并求其数学期望(注意a_t是白噪声)可得

$$\gamma_k(au) = h_k\sigma_a^2$$

所以有

$$h_k = \frac{1}{\sigma_a^2}\gamma_k(au) = \frac{\sigma_u}{\sigma_a}\rho_k(au), \qquad k \geqslant 0 \qquad (8.3.3)$$

此式及其上述的分析，是我们进行系统识别的重要依据．

依据模型(8.3.1)式进行识别估计，要利用样本数据(x_j, z_j) $(1 \leqslant j \leqslant N)$；同样，依据模型(8.3.2)式进行识别估计，要利用样本数据(a_j, u_j) $(1 \leqslant j \leqslant N)$．后者虽然由于输入化成白噪声而变得简单了，可是我们原来只有数据(x_j, z_j) $(1 \leqslant j \leqslant N)$，怎样才能获得样本$(a_j, u_j)$ $(1 \leqslant j \leqslant N)$呢？由于$x_t$的模型和参数已经确定，即$\varphi(B)\theta(B)$已知；$x_t$的样本$x_j$ $(1 \leqslant j \leqslant N)$也已获得，按照第四章或第六章的方法，可以求出$a_t(=\varphi(B)\theta^{-1}(B)x_t)$的样本估计值$\hat{a}_j(\boldsymbol{\beta}, \boldsymbol{x}_N)$，其中$\boldsymbol{\beta}$为$x_t$的模型参数向量，$\boldsymbol{x}_N$为$x_t$的样本数据向量．当$N$很大时，除前面($j$较小时)的几个$\hat{a}_j(\boldsymbol{\beta}, \boldsymbol{x}_N)$

以外，其余的样本估计值都非常接近于与样本向量 $\boldsymbol{x}_N$ 相应的样本值 a_j. 因此，我们不妨就以它们近似当做 a_j，并且在符号上也不再加区别，就以 a_j 记之. 又由于 $u_t = \varphi(B)\theta^{-1}(B)z_t$，用完全类似的方法可计算出相应于 $z_j(1 \leqslant j \leqslant N)$ 的样本估计值 $\hat{u}_j(\boldsymbol{\beta}, \boldsymbol{z}_N)$ $(1 \leqslant j \leqslant N)$. 同样当$N$很大时，它们与真正样本值 u_j 相近，因此也可以近似当做真值，并为简单起见仍用 u_j 记之. 总之，当已给数据 (x_j, z_j) $(1 \leqslant j \leqslant N)$ 后，按第四或第六章的方法便可求出相应的(近似)样本值 (a_j, u_j) $(1 \leqslant j \leqslant N)$. 利用这些数值，按照 §2 的公式可计算出样本互相关函数 $\hat{\rho}_k(a, u)$ 及 $\hat{\sigma}_u^2$ 和 $\hat{\sigma}_a^2$. 再根据(8.3.3)式，

$$\hat{h}_k = \frac{\hat{\sigma}_u}{\hat{\sigma}_a}\hat{\rho}_k(au), \qquad k \geqslant 0 \tag{8.3.4}$$

即可作为 h_k 的初步估计. 在有了系统脉冲响应函数的估计值 $\hat{h}_k$ 以后，便可依据(8.1.9)式所表明的 h_t 的诸性质进行系统的识别和估计. 下面分别对无输出干扰和有输出干扰两种情况加以讨论.

在无输出干扰时，(8.3.2)式应为

$$u_t = h(B)a_t = \zeta^{-1}(B)\Omega(B)a_{t-b} \tag{8.3.5}$$

这时，若将 $\omega_0 a_{t-b}$ 看做一个新的白噪声序列 ξ_t，那么(8.3.5)式恰好说明 u_t 相对白噪声 ξ_t 而言是一个 ARMA(r, s) 序列. 于是，按照前几章的方法，只需利用数值 $u_j(1 \leqslant j \leqslant N)$ 便可直接对 $\zeta(B)$ 和 $\frac{1}{\omega_0}\Omega(B)$ (相当于 $\varphi(B)$ 和 $\theta(B)$) 进行识别和估计. 为此，我们还总假定 $\frac{1}{\omega_0}\Omega(B)$ 满足可逆性条件. 这样识别估计所得到的 u_t 的模型记做

$$\hat{\varphi}_u(B)u_t = \hat{\theta}_u(B)\xi_t$$

其中足标 u 表示与 u_t 序列的模型有关的函数或变量. 其它类似的足标有相同含义. 那么，我们可取 $\zeta(B)$, $\frac{1}{\omega_0}\Omega(B)$ 和 $(\omega_0\sigma_a)^2$ 的估计分别为

$$\hat{\zeta}(B)=\hat{\varphi}_u(B),\quad \frac{1}{\hat{\omega}_0}\hat{\Omega}(B)=\hat{\theta}_u(B),\quad (\hat{\omega}_0\hat{\sigma}_a)^2=\hat{\sigma}_\xi^2$$

由此三式可以解出 $\hat{\zeta}_i=\hat{\varphi}_{ui}\,(1\leqslant i\leqslant p_u)$，$|\hat{\omega}_0|=\dfrac{\hat{\sigma}_\xi}{\hat{\sigma}_a}$，$\hat{\omega}_j=\hat{\omega}_0\hat{\theta}_{uj}$ $(1\leqslant j\leqslant q_u)$，$r=p_u$，$s=q_u$；其中 $\hat{\sigma}_a^2=\hat{\gamma}_0(aa)$ 可由 $a_j(1\leqslant j\leqslant N)$ 的样本方差求得．由此可见，为了完全确定出(8.3.5)式的模型估计，还须对 ω_0 的符号和 b 之值做出判断和估计．根据(8.1.9)式和(8.3.3)式，当 $k<b$ 时，应有 $\rho_k(au)=0$．又由(8.2.12)式可知，$\hat{\rho}_k(au)\,(k<b)$ 的均方根近似为 $\dfrac{1}{\sqrt{N}}$．根据这些性质，检查 $|\hat{\rho}_k(au)|\,(k=0,1,2,\cdots)$，从 k 等于什么开始，它显著地超过 $\dfrac{1}{\sqrt{N}}$（比如超过$\dfrac{2}{\sqrt{N}}$或 $\dfrac{3}{\sqrt{N}}$ 认为是显著的），我们就以这个 k 作为 b 的估计值，仍以 b 表示之．再由(8.1.9)式和(8.3.3)式知，$\gamma_b(au)=h_b\sigma_a^2$，$h_b=\omega_0$；因此可以用 $\hat{h}_b$ 的正负号来确定 $\hat{\omega}_0$ 的正负号．到此为止，便得到了(8.3.5)式的一种初步估计结果．

在有输出干扰项的情况下，由假设已知(8.3.2)式中 $h(B)a_t$ 与 ε_t 两项是相互独立的序列，因此它的识别估计问题与第六章的序列分解问题有些类似．但有利的条件是，这里第一项 $h(B)a_t$ 中的 a_t 的样本值可以由输入值 $x_j\,(1\leqslant j\leqslant N)$ 获得估计．下面将仔细分步叙述有干扰时系统的初步识别和估计方法．

1. (r,s,b) 的识别方法．b 的识别方法与无干扰时完全一样．为了识别 r 是否为零，利用 $r=0$ 时 $h_t=0\ (t\geqslant b+s+1)$ 的截尾性质．于是，可根据 $\hat{h}_{k+b+1}$ 的绝对值是否从某个 k_0 以后小于 $\dfrac{2}{\sqrt{N}}$ 或 $\dfrac{3}{\sqrt{N}}$ 判断 r 为零或不为零．若判断 $r=0$，则 k_0 即被识别为 s；若判断 $r\neq 0$，则 (r,s) 的识别是比较困难的．一般只能根据(8.1.9)式后面关于 h_t 的性质的讨论，从低阶开始逐个进行尝试．所以只有 (r,s) 不太大时才能用这个方法．

例如，x_t已定为如下的模型：

$$(1-1.97B+1.37B^2-0.34B^3)x_t=a_t$$

则

$$u_t=(1-1.97B+1.37B^2-0.34B^3)z_t$$

根据296对(x_j, z_j)的样本值计算结果为

$\hat{\sigma}_a^2=(0.188)^2$

$\hat{\sigma}_u^2=(0.358)^2$

$k=$	0	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k(au)=$	−0.01	0.05	−0.03	−0.27	−0.32	−0.43	−0.27	−0.17	−0.03	0.03	−0.05
$\hat{h}_k=$	−0.02	0.10	−0.06	−0.53	−0.63	−0.88	−0.53	−0.33	−0.06	0.06	−0.10

注意 $\frac{1}{\sqrt{N}}=\frac{1}{\sqrt{296}}\cong 0.06$；由以上数据可知，$\hat{\rho}_0(au)$，$\hat{\rho}_1(au)$，$\hat{\rho}_2(au)$ 的绝对值都不比0.06明显大，而 $|\hat{\rho}_3(au)|=0.27$ 则显著地大于0.06，因此取 $b=3$．又因 $\hat{h}_4$，$\hat{h}_5$，$\cdots$ 诸量的绝对值较大，所以判断 $r\neq 0$．但是可以认为 h_k 从 $k=5$ 开始按负指数规律衰减，此事实说明 s 可取为 $5-b=5-3=2$．至于 r 则只能从1开始逐个尝试．在 $\hat{h}_k$ $(k\geqslant 5)$ 衰减较快时，一般 r 只取1或2等不太大的值．

2. 传递部分的参数矩估计方法．当 (r, s, b) 被识别出后，在(8.1.9)式中将 h_t 代之以由(8.3.4)式算出的 $\hat{h}_t$，由此便可解出 $\zeta(B)$，$\Omega(B)$ 的参数矩估计值，即由方程组

$$\begin{cases}\hat{h}_b=\hat{\omega}_0, & k=b\\ \hat{h}_k=\hat{\zeta}_1\hat{h}_{k-1}+\hat{\zeta}_2\hat{h}_{k-2}+\cdots+\hat{\zeta}_r\hat{h}_{k-r}-\hat{\omega}_{k-b} & \\ \qquad b+1\leqslant k\leqslant b+s & \\ \hat{h}_k=\hat{\zeta}_1\hat{h}_{k-1}+\hat{\zeta}_2\hat{h}_{k-2}+\cdots+\hat{\zeta}_r\hat{h}_{k-r} & \\ \qquad b+s+1\leqslant k\leqslant b+s+r & \end{cases} \tag{8.3.6}$$

解出 $\hat{\zeta}_i$ $(1\leqslant i\leqslant r)$ 和 $\hat{\omega}_j$ $(0\leqslant j\leqslant s)$ 做为系统(8.3.1)式的传递部分参数的初步估计．

例如，继续讨论前面的例子，其阶数 (r, s, b) 已被定为(1,

2, 3)或(2, 2, 3). 若取 $r=1$, 由(8.3.6)式可得

$$\begin{cases}\hat{h}_3=\hat{\omega}_0\\ \hat{h}_4=\hat{\zeta}_1\hat{h}_3-\hat{\omega}_1\\ \hat{h}_5=\hat{\zeta}_1\hat{h}_4-\hat{\omega}_2\\ \hat{h}_6=\hat{\zeta}_1\hat{h}_5\end{cases}$$

以 $\hat{h}_k$ 的具体值代入上式解得 $\hat{\zeta}_1=0.60$, $\hat{\omega}_0=-0.53$, $\hat{\omega}_1=0.31$, $\hat{\omega}_2=0.50$; 若取 $r=2$, 用类似的方法可以求得参数的矩估计值为 $\hat{\zeta}_1=0.57$, $\hat{\zeta}_2=0.02$, $\hat{\omega}_0=-0.53$, $\hat{\omega}_1=0.33$, $\hat{\omega}_2=0.51$. 由于 $\hat{\zeta}_2=0.02\cong 0$, 所以可采用(1, 2, 3)阶作为模型阶数的估计,即原来输入输出系统的传递部分被估计为

$$(1-0.6B)y_t=(-0.53-0.31B-0.5B^2)x_t$$

至于初步识别估计的结果是否合适, 可以通过检验做出判断. 检验的方法将在本章 §5 中介绍.

3. 噪声部分的识别与估计方法. 在前两步识别与估计完成以后, 系统噪声部分识别与参数估计问题就不难解决了. 根据已有的假定,

$$n_t=z_t-h(B)x_t=\varphi_n^{-1}(B)\theta_n(B)a_t$$

是 ARMA 序列, 所以只须按前几章的方法对 n_t 进行模型识别和参数估计. 所用数据 $n_j\,(1\leqslant j\leqslant N)$ 可由原数据 $(x_j,\ z_j)\ (1\leqslant j\leqslant N)$ 和前两步对 $h(B)$ 的估计结果得到,即令 $\hat{h}(B)=\hat{\zeta}^{-1}(B)\hat{\omega}(B)$ 以及

$$n_j=z_j-\hat{h}(B)x_j,\qquad j=1,2,\cdots,N \tag{8.3.7}$$

其中凡用到 $x_j\,(j\leqslant 0)$ 时,可以用零代替.由 $\zeta(B)$ 的稳定性可知,当 N 很大时, 这样算出的 n_j 除前面的若干个以外,其余都与真实样本值相近,因此可把它们当做 n_t 的样本值使用. 不过, 如果只是为了求 n_t 的初步识别和矩估计,不必由(8.3.7)式计算 n_t 的样本值, 可用其它办法直接求出 ε_t 的样本自相关函数,然后进行识别和估计. 其具体方法如下:

令 $v_t=h(B)a_t$ 于是(8.3.2)式可改写成

$$u_t=v_t+\varepsilon_t$$

因为已假定 a_t 与 ε_t 相互独立，所以 v_t 与 ε_t 也相互独立．由此可得

$$\gamma_k(u) = \gamma_k(v) + \gamma_k(\varepsilon) \tag{8.3.8}$$

由 $v_t = h(B)a_t = \sum_{k=0}^{\infty} h_k a_{t-k}$ 以及 (8.3.3) 式可知，对 $k > 0$

$$\gamma_k(v) = \sigma_a^2 \sum_{j=0}^{\infty} h_j h_{j+k} = \sigma_u^2 \sum_{j=0}^{\infty} \rho_j(au)\rho_{j+k}(au)$$

以此式代人 (8.3.8) 式可得

$$\begin{aligned}\gamma_k(\varepsilon) &= \gamma_k(u) - \sigma_u^2 \sum_{j=0}^{\infty} \rho_j(au)\rho_{j+k}(au) \\ &= \sigma_u^2 \left\{\rho_k(u) - \sum_{j=0}^{\infty} \rho_j(au)\rho_{j+k}(au)\right\} \end{aligned} \tag{8.3.9}$$

故有

$$\rho_k(\varepsilon) = \left\{\rho_k(u) - \sum_{j=0}^{\infty} \rho_j(au)\rho_{j+k}(au)\right\}\left\{1 - \sum_{j=0}^{\infty} \rho_j^2(au)\right\}^{-1} \tag{8.3.10}$$

由于在"预白化"过程中已经计算过样本互相关函数 $\hat{\rho}_j(au)$和样本自相关函数 $\hat{\rho}_k(u)$，所以，在(8.3.9)，(8.3.10) 两式中将 $\rho_k(u)$，$\rho_k(au)$ 和 σ_u^2 分别代之以 $\hat{\rho}_k(u)$，$\hat{\rho}_k(au)$和 $\hat{\sigma}_u^2 = \hat{\gamma}_0(u)$，便得到 ε_j ($1 \leqslant j \leqslant N$) 的近似样本自协方差函数 $\hat{\gamma}_k(\varepsilon)$ 和样本自相关函数 $\hat{\rho}_k(\varepsilon)$．(8.3.9)，(8.3.10)两式中的无穷求和可以用有穷和近似．由于这些相关函数的理论值都很快衰减到零，所以这种近似是可行的．有了 ε_t 的样本自相关函数，按第二章 §2 的 (2.2.23) 式还可求出其样本偏相关函数 $\phi_{kk}(\varepsilon)$．然后，用第三章的方法便可以定出 ε_t 的模型初步估计．如果把识别估计出的模型记做

$$\hat{\varphi}_\varepsilon(B)\varepsilon_t = \hat{\theta}_\varepsilon(B)\alpha_t$$

其中 α_t 为白噪声，那么，由于 $\varepsilon_t = \varphi(B)\theta^{-1}(B)n_t$，以此代入上式即得 n_t 的估计模型：

$$\varphi(B)\hat{\varphi}_\varepsilon(B)n_t = \theta(B)\hat{\theta}_\varepsilon(B)\alpha_t \tag{8.3.11}$$

其中 $\varphi(B)$，$\theta(B)$ 是由 x_t 的模型给出的．

例如，继续分析本节前面所讨论的数值例子．由已算出的 $\hat{\sigma}_a^2$ 和 $\hat{\rho}_j(au)$ 等值及(8.3.10)式，算得 $\hat{\rho}_k(\varepsilon)$ $(k=0, 1, 2, \cdots)$ 依次取如下的值:

$$1,\ -0.382,\ 0.126,\ -0.011,\ 0.064,\ \cdots$$

又由 $\hat{\rho}_k(\varepsilon)$ 算得 $\hat{\varphi}_{kk}(\varepsilon)$ $(k=1, 2, \cdots)$ 依次为

$$-0.382,\ -0.01,\ 0.04,\ \cdots$$

由此可判断 ε_t 为 AR(1, 0) 序列，估计结果为

$$(1+0.38B)\varepsilon_t = a_t$$

注意 x_t 的模型已定为 AR(3, 0)，将它们代入(8.3.11)式即得 n_t 的估计模型

$$(1+0.38B)(1-1.97B+1.37B^2-0.34B^3)n_t = a_t$$

即

$$(1-1.59B+0.62B^2+0.18B^3-0.13B^4)n_t = a_t$$

到此为止，对于本节的例子，我们已获得全部识别和估计的结果，现在仔细列出这个系统的初步估计模型如下:

$$z_t = (-0.53-0.31B-0.50B^2)(1-0.60B)^{-1}x_{t-3}$$
$$+(1-1.59B+0.62B^2+0.18B^3-0.13B^4)^{-1}a_t$$

§4. 系统参数的最小二乘估计

前一节介绍的系统参数估计方法，是将传递部分与噪声部分分别进行估计的，而且只讨论了矩估计方法．与第四章中 ARMA 序列参数的精估计类似，同样可以定义系统(8.3.1)式中诸参数的各种精估计；除了计算的复杂程度大大增加外，其数学原理是完全相同的．我们在这一节中，只简单介绍精估计中的最小二乘估计．

在上一节，无输出干扰噪声的(8.3.1)式的识别估计问题，已被转化成 ARMA 模型(8.3.5)式的识别估计问题．因此，可以用第四章的各种精估计方法解决这里的问题．这里不必一一重复．

对于有输出干扰噪声的系统，首先按照上一节的方法，对系统传递部分和噪声部分的阶数 (r, s, b) 和 (p_n, q_n) 进行初步识别，而且不妨就用 (r, s, b) 和 (p_n, q_n) 代表识别的结果．然后，用

$\boldsymbol{\theta}_n$, $\boldsymbol{\varphi}_n$, $\boldsymbol{\zeta}$ 和 $\boldsymbol{\Omega}$ 分别表示 $\theta_n(B)$, $\varphi_n(B)$, $\zeta(B)$ 和 $\Omega(B)$ 的系数组成的向量. 它们都是未知参数，是待估计的. 现在重新改写 (8.1.17) 式为

$$\theta_n(B)a_t = \varphi_n(B)z_t - \zeta^{-1}(B)\omega(B)\varphi_n(B)x_t$$

或

$$a_t = \theta_n^{-1}(B)\varphi_n(B)z_t - \zeta^{-1}(B)\Omega(B)\theta_n^{-1}(B)\varphi_n(B)x_{t-b} \quad (8.4.1)$$

回忆一下第四章 §1 中 ARMA 序列参数的最小二乘估计的定义，便不难仿照它来定义 (8.4.1) 式中诸参数的最小二乘估计. 在给出样本序列 (x_j, z_j) $(1 \leqslant j \leqslant N)$ 和任一组参数值以后，由(8.4.1) 式可求得相应的 a_t 的样本序列 $a_j(1 \leqslant j \leqslant N)$. 其中凡用到 x_j, z_j $(j \leqslant 0)$ 均以零值代之. 如果让参数值变化，这样算得的 a_j 就应表成参数的函数形式，即 $a_j = a_j(\boldsymbol{\theta}_n, \boldsymbol{\varphi}_n, \boldsymbol{\zeta}, \boldsymbol{\Omega})$. 在 $\boldsymbol{\theta}_n$, $\boldsymbol{\varphi}_n$, $\boldsymbol{\zeta}$ 和 $\boldsymbol{\Omega}$ 的所有可能值中选出 $\hat{\boldsymbol{\theta}}_n$, $\hat{\boldsymbol{\varphi}}_n$, $\hat{\boldsymbol{\zeta}}$, $\hat{\boldsymbol{\Omega}}$, 使

$$S(\boldsymbol{\theta}_n, \boldsymbol{\varphi}_n, \boldsymbol{\zeta}, \boldsymbol{\Omega}) = \sum_{j=1}^{N} a_j^2(\boldsymbol{\theta}_n, \boldsymbol{\varphi}_n, \boldsymbol{\zeta}, \boldsymbol{\Omega}) \quad (8.4.2)$$

达到极小，$\hat{\boldsymbol{\theta}}_n$, $\hat{\boldsymbol{\varphi}}_n$, $\hat{\boldsymbol{\zeta}}$, $\hat{\boldsymbol{\Omega}}$ 就是诸参数的最小二乘估计；而

$$\hat{\sigma}_a^2 = \frac{1}{N} S(\hat{\boldsymbol{\theta}}_n, \hat{\boldsymbol{\varphi}}_n, \hat{\boldsymbol{\zeta}}, \hat{\boldsymbol{\Omega}})$$

就是白噪声 a_t 的方差 σ_a^2 的估计.

由 (8.4.1) 式计算 a_j 时，可以用直接求无穷级数和的办法，如

$$\theta_n^{-1}(B)\varphi_n(B)z_j = \psi_n(B)z_j = \sum_{i=0}^{\infty} \psi_{ni} z_{j-i} \cong \sum_{i=0}^{j-1} \psi_{ni} z_{j-i}$$

其中 ψ_{ni} 是序列 n_t 的传递函数. 但是，这种算法的计算量太大. 一般都用递推公式求 a_j，即由 (8.4.1) 式导出

$$\zeta(B)\theta_n(B)a_t = \varphi_n(B)\zeta(B)z_t - \Omega(B)\varphi_n(B)x_{t-b} \quad (8.4.3)$$

再由此式改写成

$$a_t = \sum_{i=1}^{r+q_n} \mu_i a_{t-i} + \sum_{i=0}^{r+p_n} v_i z_{t-i} - \sum_{i=0}^{s+p_n} \eta_i x_{t-b-i} \quad (8.4.4)$$

其中 μ_i, v_i 和 η_i 是由 (8.4.3) 式中 (给定的) 诸参数所决定的系数，即

$$\zeta(B)\theta_n(B) = 1 - \sum_{i=1}^{r+q_n} \mu_i B^i, \quad \varphi_n(B)\zeta(B) = \sum_{i=0}^{r+p_n} \nu_i B^i$$

$$\Omega(B)\varphi_n(B) = \sum_{i=0}^{r+p_n} \eta_i B^i$$

在 (8.4.4) 式中取 $t = 1, 2, \cdots, N$, 以及初值

$$\alpha_j = z_j = x_j = 0, \qquad j \leqslant 0$$

便可递推算得 $\alpha_1, \alpha_2, \cdots, \alpha_N$; 它们都是未知参数的函数，为了简便，只用 α_j 表示.

至于如何求解 (8.4.2) 式的极小值，则是计算方法问题. 第四章§3 所介绍的方法在这里完全适用，我们不再一一重述. 此外，读者还可以在现代文献中寻求更合适的求极值方法，或者针对具体问题采取一些灵活的技巧以减少计算量.

如果数据量 N 不够大，为了提高估计的精度，还可以采用一种更细致的最小二乘估计方法. 首先注意，在用 (8.4.4) 式计算 α_j 时，由于人为地将所有初值 α_j, z_j, x_j ($j \leqslant 0$) 都取为零，当 j 较小时，α_j 显然受初值影响很大. 为了减少这种人为地加入初始值的影响，一方面可将 (8.4.2) 式的平方和改用

$$S(\theta_n, \varphi_n, \zeta, \Omega) = \sum_{j=m+1}^{N} \alpha_j^2 \tag{8.4.5}$$

其中 $m = \max(r + p_n, b + s + p_n)$. 另一方面，将零初值改用一组估计值代替. 事实上，由于 (8.4.5) 式中的 $j \geqslant m + 1$，所以在用 (8.4.4) 式递推计算 $\alpha_j (j \geqslant m + 1)$ 时. 并不涉及初值 z_j, x_j ($j \leqslant 0$)，而只用到 $\alpha_m, \alpha_{m-1}, \cdots, \alpha_{m-r-q_n+1}$; 它们可用线性最小二乘估计值来代替. 具体做法是，在每给定一组参数 ($\theta_n, \varphi_n, \zeta, \Omega$) 时，把 $\alpha_m, \alpha_{m-1}, \cdots, \alpha_{m-r-q_n+1}$ 当做自由变量; 在 (8.4.4) 式中取 $t = m + 1, m + 2, \cdots, m + k_0$ (k_0 为适当选定的比 $r + q_n$ 大，比 N 小的正整数)，得到 k_0 个 α_j ($m+1 \leqslant j \leqslant m + k_0$) 关于这些自由变量的表达式; 在这些表达式中，涉及到的 x_j, z_j ($1 \leqslant j \leqslant m + k_0$) 都是已知数据，因此可将它们表成

$$\alpha_j = f_j - (a_{j0}\alpha_m + a_{j1}\alpha_{m-1} + \cdots + a_{j\tau}\alpha_{m-\tau})$$
$$m + 1 \leqslant j \leqslant m + k_0$$

其中 $\tau = r + q_n - 1$，f_j 是数据 $(x_j, z_j)(1 \leqslant j \leqslant m + k_0)$ 的线性函数，a_{ji} 是系统参数的函数．上式可改写成

$$f_j = a_{j0}\alpha_m + a_{j1}\alpha_{m-1} + \cdots + a_{j\tau}\alpha_{m-\tau} + \alpha_j$$
$$m + 1 \leqslant j \leqslant m + k_0 \tag{8.4.6}$$

再添上恒等式组

$$0 = -\alpha_i + \alpha_i, \qquad m - \tau \leqslant i \leqslant m \tag{8.4.7}$$

(8.4.6)和(8.4.7)式恰好构成第一章 §5 中提到的线性最小二乘模型，由此即可给出自由变量 $\alpha_m, \alpha_{m-1}, \cdots, \alpha_{m-\tau}$ 的最小二乘估计(用(1.5.3)式)．这就是刚刚说过的用估值代替零初值时关于 $\alpha_j(m-\tau \leqslant j \leqslant m)$ 的算法．然后，再用(8.4.4)式求出其余的 $\alpha_j(m+1 \leqslant j \leqslant N)$，继而由(8.4.5)式算出 $S(\boldsymbol{\theta}_n, \boldsymbol{\varphi}_n, \boldsymbol{\zeta}, \boldsymbol{\Omega})$．在所有容许的参数值中，以使 S 达极小值的参数做为其估计．这种估计比上面的最小二乘估计要稍好一些；但是增加了计算内容，而且关于(8.4.6)式仔细形式的获得办法尚待讨论．自然，我们可以根据系统的模型和(8.4.6)式中 f_j，a_{ji} 的含义，给出它们的一般表达式，但那将是十分烦琐和冗长的．因此，我们仅通过简单的低阶模型的例子，说明建立(8.4.6)式的具体形式的方法，以使读者掌握其要领．

例如，经初步识别后，已确定某系统的阶数估计分别为 $r=1$，$s=0$，$b=1$，$p_n=0$，$q_n=1$．我们要给出以下模型中诸参数的细致的最小二乘估计，即

$$(1-\zeta B)(1-\theta B)\alpha_t = (1-\zeta B)z_t - \omega_0 x_{t-1} \tag{8.4.8}$$

在此例中，$m = \max(r + p_n, b + s + p_n) = 1$，$\tau = r + q_n - 1 = 1$，

$$S(\theta, \zeta, \omega_0) = \sum_{j=2}^{N} \alpha_j^2 \tag{8.4.9}$$

初始值变量为 α_0，α_1．由(8.4.8)式得

$$\alpha_j = (\zeta + \theta)\alpha_{j-1} - \zeta\theta\alpha_{j-2} + z_j - \zeta z_{j-1} - \omega_0 x_{j-1} \quad (8.4.10)$$

这个式子有两重用处．其一，令 $j = 2, 3, \cdots, k_0 + 1$，以求出 α_0, α_1 的最小二乘估计，做为以后递推 α_j 的初值．其二，给出 (8.4.9) 式中诸 α_j 的递推计算值．后一用处是简单而明了的，而前一用处则还须进一步做些说明．由 (8.4.10) 式可令

$$\alpha_j = -(a_{j0}\alpha_1 + a_{j1}\alpha_0) + f_j, \quad 2 \leqslant j \leqslant k_0 + 1 \quad (8.4.11)$$

再令

$$g_j = z_j - \zeta z_{j-1} - \omega_0 x_{j-1}$$

则有

$$\begin{aligned}-(a_{j0}\alpha_1 + a_{j1}\alpha_0) + f_j = \alpha_j &= (\zeta + \theta)\alpha_{j-1} - \zeta\theta\alpha_{j-2} + g_j \\ &= -(\zeta + \theta)(a_{j-1,0}\alpha_1 + a_{j-1,1}\alpha_0 - f_{j-1}) + \zeta\theta(a_{j-2,0}\alpha_1 \\ &\quad + a_{j-2,1}\alpha_0 - f_{j-2}) + g_j = -[(\zeta + \theta)a_{j-1,0} - \zeta\theta a_{j-2,0}]\alpha_1 \\ &\quad - [(\zeta + \theta)a_{j-1,1} - \zeta\theta a_{j-2,1}]\alpha_0 + (\zeta + \theta)f_{j-1} \\ &\quad - \zeta\theta f_{j-2} + g_j\end{aligned}$$

通过对比系数可以得到 f_j, a_{ji} 的递推计算公式如下：

$$\begin{cases}a_{j0} = (\zeta + \theta)a_{j-1,0} - \zeta\theta a_{j-2,0}, & 2 \leqslant j \leqslant k_0 + 1 \\ a_{j1} = (\zeta + \theta)a_{j-1,1} - \zeta\theta a_{j-2,1}, & 2 \leqslant j \leqslant k_0 + 1 \\ f_j = (\zeta + \theta)f_{j-1} - \zeta\theta f_{j-2} + g_j, & 2 \leqslant j \leqslant k_0 + 1\end{cases}$$

其中初值是明显的．事实上，由于

$$\alpha_1 = -(a_{1,0}\alpha_1 + a_{1,1}\alpha_0) + f_1, \quad \alpha_0 = -(a_{0,0}\alpha_1 + a_{0,1}\alpha_0) + f_0$$

所以初值应是：$a_{1,0} = -1$，$a_{1,1} = 0$，$a_{0,0} = 0$，$a_{0,1} = -1$，$f_0 = f_1 = 0$．

为了进一步熟悉建立 (8.4.6) 式的方法，我们给出上例的某些数值结果．即在参数值取 $\zeta = 0.5$，$\theta = 0.2$，$\omega_0 = 0.1$ 时计算出 a_{ji} 和 f_j．先将这些参数代入 a_{ji} 和 f_j 的递推表达式得

$$\begin{cases}a_{j,0} = 0.7a_{j-1,0} - 0.1a_{j-2,0}, & 2 \leqslant j \leqslant 8 \\ a_{j,1} = 0.7a_{j-1,1} - 0.1a_{j-2,1}, & 2 \leqslant j \leqslant 8 \\ f_j = 0.7f_{j-1} - 0.1f_{j-2} + g_j, & 2 \leqslant j \leqslant 8\end{cases}$$

这里取 $k_0 = 7$．由 (8.4.6) 与 (8.4.7) 式，α_0 和 α_1 的线性最小二乘估计的模型为

表 8.4.1 (8.4.7)式中 α_i 的样本值 (参数为 $\zeta = 0.7$，$\theta = 0.2$，$\omega_0 = 0.1$)

j	x_j	z_j	$g_j = z_j - 0.5z_{j-1} - 0.1x_{j-1}$	$f_j = 0.7f_{j-1} - 0.1f_{j-2} + g_j$	$a_{j,0} = 0.7a_{j-1,0} - 0.1a_{j-2,0}$	$a_{j,1} = 0.7a_{j-1,1} - 0.1a_{j-2,1}$	$f_j = a_{j,0}\alpha_1 + a_{j,1}\alpha_0 + \alpha_j$
0				0	0	−1	$0 = -\alpha_0 + \alpha_0$
1	30	−2		0	−1	0	$0 = -\alpha_1 + \alpha_1$
2	0	2	0	0	−0.7	0.1	$0 = -0.7\alpha_1 + 0.1\alpha_0 + \alpha_2$
3	−10	2	1	1	−0.39	0.07	$1 = -0.39\alpha_1 + 0.07\alpha_0 + \alpha_3$
4	0	0	0	0.7	−0.203	0.039	$0.7 = -0.203\alpha_1 + 0.039\alpha_0 + \alpha_4$
5	−40	4	4	4.39	−0.103	0.020	$4.39 = -0.103\alpha_1 + 0.02\alpha_0 + \alpha_5$
6	0	−2	0	3.00	−0.052	0.010	$3.00 = -0.052\alpha_1 + 0.01\alpha_0 + \alpha_6$
7	−10	0	1	2.66	−0.026	0.005	$2.66 = -0.026\alpha_1 + 0.005\alpha_0 + \alpha_7$
8	10	−2	−1	0.56	−0.013	0.003	$0.56 = -0.013\alpha_1 + 0.003\alpha_0 + \alpha_8$

$$\begin{cases} f_j = a_{j,0}\alpha_1 + a_{j,1}\alpha_0 + \alpha_j, & 2 \leqslant j \leqslant 8 \\ 0 = -\alpha_1 + \alpha_1 \\ 0 = -\alpha_0 + \alpha_0 \end{cases}$$

在表 8.4.1 中列出了具体数据和有关量的计算结果. 表 8.4.1 中最右边的九个方程式构成关于 α_0 和 α_1 的最小二乘模型; 用 (1.5.3) 式的方法可得

$$\begin{pmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_0 \end{pmatrix} = (A^\tau A)^{-1} A^\tau F = \begin{pmatrix} 1.697 & -0.108 \\ -0.108 & 1.017 \end{pmatrix}^{-1} \begin{pmatrix} -1.188 \\ 0.230 \end{pmatrix}$$

$$= \begin{pmatrix} -0.69 \\ 0.15 \end{pmatrix}$$

其中 A 是 (9×2) 阶的矩阵, 其元素为 $a_{ji}(0 \leqslant j \leqslant 8, 0 \leqslant i \leqslant 1)$; F 为 9 维向量, 其元素为 f_j $(0 \leqslant j \leqslant 8)$. 再以此做为 α_j 的初值, 即 $\alpha_1 = -0.69$, $\alpha_0 = 0.15$, 代入 (8.4.10) 式即可递推求得

$$\alpha_2 = -0.50, \quad \alpha_3 = 0.72, \quad \alpha_4 = 0.55, \quad \cdots$$

至于如何求解 (8.4.5) 式的极小值问题, 同样可用第四章 §3 的有关方法, 或者选用其它数值解法, 这里不去详细介绍了.

§5. 系统模型的检验

在系统模型的识别与参数估计被获得以后, 为了鉴定估计模型的可信程度, 还须对它进行检验.

对无输出干扰项的系统检验, 等价于对模型 (8.3.5) 式的检验. 而 (8.3.5) 式中的 u_t 为 ARMA 序列, 因此, 第五章所介绍的各种检验方法便可在此使用.

对有输出干扰的系统 (8.1.17) 的检验, 等价于对模型 (8.1.18) 式的检验. 在这一节里, 如同在第五章一样, 尽管 (8.1.17) 式的参数可能都是估计值 (在个别情况下, 也可能通过其它手段获得模型参数), 我们并不用上边加 "^" 号的字母表示, 而仍用 ($\boldsymbol{\theta}_n$, $\boldsymbol{\varphi}_n$, ζ, $\boldsymbol{\Omega}$, σ_a^2) 表示. 这不仅是为了符号的简练, 还由于本书所说的检验既用于检验估计模型, 也可用于检测模型是否发生变化.

现在, 假定有一组样本数据 (x_j, z_j) $(1 \leqslant j \leqslant N)$, 我们要检

验它们是否为系统(8.1.17)式的输入输出的样本值，或者说，要检验它们能否用(8.1.17)式的模型进行拟合. 在上一节中，已从(8.1.17)式导出(8.4.1)和(8.4.3)式，从而导出(8.4.4)式，即

$$\alpha_j = \sum_{i=1}^{r+q_n} \mu_i \alpha_{j-i} + \sum_{i=0}^{r+p_n} \nu_i z_{j-i} - \sum_{i=0}^{s+p_n} \eta_i x_{j-b-i} \qquad (8.5.1)$$

为了从样本值(x_j, z_j) $(1 \leqslant j \leqslant N)$获得相应的$\alpha_t$的样本估计值$\alpha_j(1 \leqslant j \leqslant N)$，可由(8.5.1)式进行递推计算. 但是，其中还要用到一些初值量 $z_j(-r-p_n+1 \leqslant j \leqslant 0)$，$x_j(-b-s-p_n+1 \leqslant j \leqslant 0)$ 和 $\alpha_j(-r-q_n+1 \leqslant j \leqslant 0)$. 对于这些初始值的选取，有各种不同的方法. 最简单的方法是都取零值(即它们的均值)；比较合理的方法是取它们的估计值，即在已给数据(x_j, z_j) $(1 \leqslant j \leqslant N)$时的线性最小方差估计值. 界于上述两种方法之间的办法，是上节所介绍的用线性最小二乘法给出初值的估计. 以上三种方法，除第二种外在上一节都已介绍过. 在应用中，主要用一、三两种方法，因此这里不去仔细介绍第二种方法了. 无论用哪种方法，在获得了初值以后，便可从(8.5.1)式求出相应的样本值$\alpha_j(1 \leqslant j \leqslant N)$. 不过，如果用最小二乘法给出$\alpha_j$的初始值(具体方法见前一节的叙述)，则只能获得 $m+1 \leqslant j \leqslant N$ 之间的 α_j $(m = \max(r+p_n, b+s+p_n))$. 这一组值受初值的影响较小，因而更近于它们的真实样本值.

根据系统(8.1.17)式的定义，α_t是x_t的模型里的白噪声，因此，可以用第五章所介绍的方法，检验上面所计算出的样本值 $\alpha_1, \alpha_2, \cdots, \alpha_N$ (或 $\alpha_{m+1}, \alpha_{m+2}, \cdots, \alpha_N$)，是否为白噪声序列的样本值. 具体方法这里不重复了.

如果经检验可以确认"这组α_j是白噪声的样本值"，便认为模型(8.1.17)式可以被接受，或者说这个模型对于这组数据拟合得较好. 如果经检验否定"这组α_j是白噪声的样本值"，那么，一般就认为模型(8.1.17)式不能被接受，或者说这个模型对于所给的数据拟合得不好.

但是，当这样检验后得到否定结果时，我们还不能找出其原

因．为了进一步细致分析原模型为什么不能被接受，可以对 α_t 与 x_t 是否为不相关的序列进行假设检验．由于 x_t 的模型 $\varphi(B)x_t=\theta(B)a_t$ 是已知的，所以可按第五章的办法求出相应于样本 $x_j(1\leqslant j\leqslant N)$ 的 a_t 的样本值 $a_j(1\leqslant j\leqslant N)$，且仍用 a_j 表示．这样，我们就可以检验假设"a_j $(1\leqslant j\leqslant N)$ 是白噪声 a_t 的样本值；白噪声 a_t 与 α_t 是互不相关的"．具体检验方法是，先按(8.2.5)和(8.2.6)式求出 a_t 与 α_t 的样本互相关函数 $\hat{\rho}_k(a\alpha)$．如果原假设成立，根据(8.2.13)式中 $\hat{\rho}_k(a\alpha)$ 的渐近性质可知，当 $N\to\infty$ 时，$\hat{\rho}_k(a\alpha)$ 为渐近正态分布 $N\left(0,\frac{1}{\sqrt{N}}\right)$，而且 $\hat{\rho}_k(a\alpha)$ 与 $\hat{\rho}_j(a\alpha)$（当 $k\neq j$ 时）渐近独立．因此若取

$$Q_K=N\sum_{j=1}^{K}[\hat{\rho}_j^2(a\alpha)+\hat{\rho}_{-j}^2(a\alpha)]$$

则 Q_K 的分布渐近于自由度为 $2K$ 的中心 χ^2 分布．用 χ^2 检验可以对原假设做出是否被接受的判断．如果经上述检验后可接受原假设，而前面对 α_j 是白噪声的假设检验被否定，结合两次检验结果，可以认为原模型(8.1.17)式的传递部分拟合尚好．对于其噪声部分的模型估计可进行必要的修改，以寻求更合适的模型估计(或拟合)．其修改方法可以参考第六章的方法．如果经上述检验后否定原假设，则要重新修改(8.1.17)式模型，甚至可能真实系统不适于用本章的有限阶系统描述．

最后，我们还须补充一点说明．在本章的前几节都假定系统输入 x_t 和输出 z_t 的均值为零，但是真实情况往往并不如此．当然，进行差分可以使 $\nabla^D z_t$ 和 $\nabla^d x_t$ 的均值近似为零；但是，当我们有较精确的手段获得 Ex_t 或 Ez_t 时，可以用 (x_j-Ex_j) 和 (z_j-Ez_j) 代替 x_j 和 z_j $(1\leqslant j\leqslant N)$ 进行系统识别．下面仅就 Ex_t 为已知常数时的情况加以讨论．

在(8.1.17)式中，仍假定 α_t 是与 x_t 独立的均值为零的白噪声，而 (x_t-Ex_t) 为 ARMA（或 ARIMA）序列．记 $Ex_t=\mu$，$Ez_t=\nu_t$，那么，在(8.1.17)式两边取均值可得

$$\nu_t = Ez_t = E\zeta^{-1}(B)\omega(B)x_t = Eh(B)x_t = \sum_{j=1}^{\infty} h_j\mu = h(1)\mu$$

其中

$$h(1) = \zeta^{-1}(1)\omega(1) = \frac{\omega_0 - \omega_1 - \cdots - \omega_s}{1 - \zeta_1 - \cdots - \zeta_r}$$

称为系统的稳态增益. 于是有

$$\nu = \nu_t = Ez_t = \frac{\omega_0 - \omega_1 - \cdots - \omega_s}{1 - \zeta_1 - \cdots - \zeta_r}\mu \qquad (8.5.2)$$

从表面上看，在依据 (x_j, z_j) $(1 \leqslant j \leqslant N)$ 进行系统识别时，似乎可以用均值为零的样本 $(x_j - \mu, z_j - \nu)$ $(1 \leqslant j \leqslant N)$ 代替原来的样本，以便采用本章的识别方法. 但是，在只知道 x_t 的均值常数 μ 时，z_t 的均值常数 ν 却依赖于未知参数 $\omega_0, \omega_1, \cdots, \omega_s$ 和 $\zeta_1, \zeta_2, \cdots, \zeta_r$，因此识别的手续就更复杂了. 通常只能用近似方法处理，即令

$$\hat{\mu} = \frac{1}{N}\sum_{j=1}^{N} x_j, \qquad \hat{\nu} = \frac{1}{N}\sum_{j=1}^{N} z_j$$

而把 $(x_j - \hat{\mu}, z_j - \hat{\nu})$ $(1 \leqslant j \leqslant N)$ 近似当作均值为零的样本进行识别和估计，并将其结果做为对 (x_t, z_t) 系统的拟合描述. 反过来还可以检查 $\hat{\zeta}^{-1}(1)\hat{\omega}(1)\hat{\mu}$ 是否与 $\hat{\nu}$ 近似相等，若相差很大，说明系统拟合得不好.

在输入量 x_t 的均值 $Ex_t = \mu_t$ 不为常值时，系统识别就会更加困难. 一般说来，采用求和模型(即把 x_t 当作 ARIMA 序列)的处理方法还算是比较简单的.

第九章　随机漂移的控制

§1. 概　　述

在绪论中我们曾举过轧钢过程的例子．为了使轧制后的钢板具有预定的厚度 h'，理论上需要把轧辊的辊缝固定为一定的量值 μ，但是由于实际输入的钢板厚度 h_k 在均值 h 上下起伏，因此在设计轧钢自动控制系统时，就应当考虑如何根据 h_k 的变化而适当调整辊缝，即围绕其均值 μ 加一调剂量 x_k，使轧制后的钢板厚度 h'_k 在预定值 h' 上下的起伏量 $y_k = h'_k - h'$ 尽量小．这类问题就属于随机漂移控制的范畴．又如某种化学反应器，设其输入的化学原料为甲，反应后的生成物为乙，加入的催化剂为丙．由于甲的浓度不断变化，生成物的浓度也会相应改变，为了使乙的浓度尽量稳定在所要求的量值附近，就要提出如何合理控制催化剂丙的输入量的问题．在很多实际领域中都可能遇到上述类似的控制问题．本章将要介绍的随机漂移控制方法，是解决这类问题的一种数学方法．

我们将这类实际问题概括成如下的数学问题：假定有一个多输入单输出的线性系统（见图 9.1.1 的虚线部分），其输入量分为控制量与噪声量两种．输入噪声量又分成可量测的和不可量测的两种；我们依次用 x_t，z_t和 u_t 分别表示上述三种量．又假定可量测噪声 z_t 与不可量测噪声 u_t 是相互独立的．进一步，还假定整个系统的输出量 ε_t 等于三个有限阶单输入系统输出量的简单叠加，即如果把 z_t 通过系统 II 的输出记做 y_t，u_t 通过系统 I 的输出记做 N_t，x_t 通过系统 III 的输出记做 y_t^*，则 $\varepsilon_t = y_t + N_t + y_t^*$．且这三个系统都是前一章讨论过的有限阶系统．$\varepsilon_t$ 称为整个系统的输出漂移．所谓最佳输出漂移控制，是指设计一个馈给系统 IV（见图 9.1.1 曲线部分），它以可量测噪声 z_t 和漂移量 ε_t 为其输入，以

x_t 为其输出，用这样的 x_t 作为控制量，以使 ε_t 的方差达到极小值.

在所举的轧钢的例子中，输入钢板的厚度起伏量 $z_k = h_k - h$ 是可量测的噪声，辊缝调剂量 x_k 是控制量，而输入钢板的硬度的

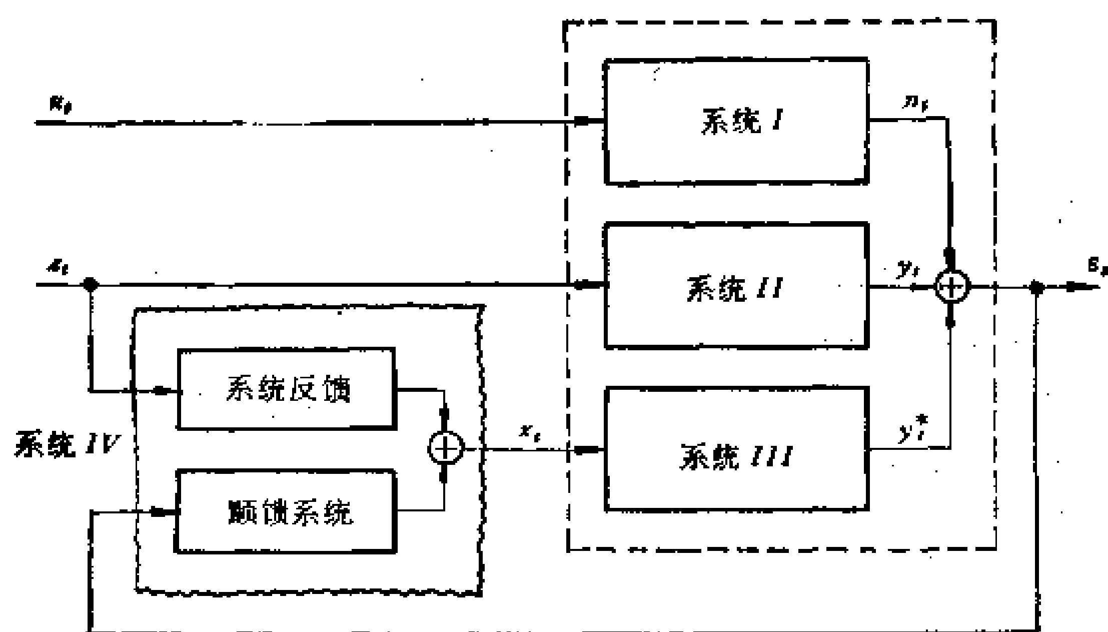


图 9.1.1 随机漂移控制示意图

起伏量一般是不可量测的噪声. 这两种噪声与输出的关系分别用系统 II 和系统 I 来描述，调制量与输出的关系用系统 III 描述. 在绪论和第八章中我们曾分别论述过这些系统的识别问题和识别方法. 在这一章里，我们假定它们都已经被识别而确定了. 现在的任务是，进一步讨论如何根据这些系统模型来设计最佳的控制馈给系统，以尽量压低输出漂移的方差. 当然，除去上述的诸假定以外，为了使用 ARMA 序列分析的技巧，还要假定能用 ARIMA 序列来描述所有的输入和输出诸量.

从上述分析可以看出，设计 x_t 的馈给系统 IV 与系统 I 并无直接关系；因此，在最佳漂移控制设计的讨论中，我们无须写出系统 I，只须定出 N_t 的模型即可；在以后的讨论中，直接假定 N_t 的模型是给定的. 这样一来，系统 II，III 和 IV 是我们以后要讨论的主要对象. 系统 II 可以称之为工作系统，系统 III 可以称之为控制系统，系统 IV 叫做馈给系统. 以后几节主要讨论最佳馈给系统的

具体结构的设计方法.

§2. 顺馈系统

在这一节里,讨论在没有不可量测输入噪声的情况下,如何解决前一节所提出的最佳漂移控制问题. 具体地说,假定有一个双输入单输出系统(如图 9.2.1 所示),其可量测噪声输入 z_t 与相应的输出 y_t 的系统(即工作系统)关系为

$$\zeta(B)y_t = \omega(B)z_t \text{ 或 } \zeta(B)y_t = \Omega(B)B^b x_t \quad (9.2.1)$$

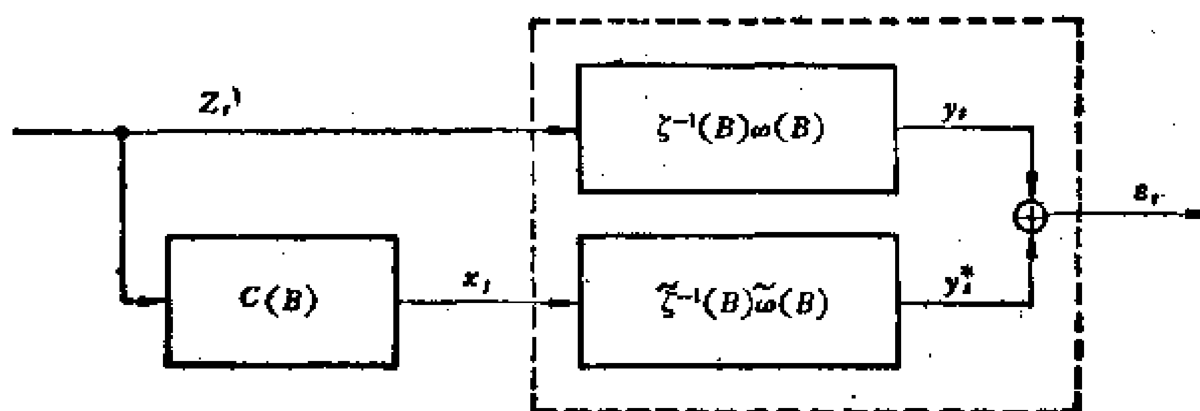


图 9.2.1 顺馈控制示意图

其控制量输入 x_t 与相应的输出 y_t^* 的系统(即控制系统)关系为

$$\tilde{\zeta}(B)y_t^* = \tilde{\omega}(B)x_t \text{ 或 } \tilde{\zeta}(B)y_t^* = \tilde{\Omega}(B)B^f x_t \quad (9.2.2)$$

以上两式中 $\zeta(B)$, $\Omega(B)$, $\tilde{\zeta}(B)$, $\tilde{\Omega}(B)$ 如同 §8.1 中的定义,都是一步延迟算子 B 的多项式,而且假定它们满足稳定性和可逆性条件,即它们的根都在单位圆外. b 和 f 分别表示系统(9.2.1)和(9.2.2)的延迟步数. 我们暂时只考虑如下形式的馈给系统

$$x_t = C(B)z_t \quad (9.2.3)$$

其中 $C(B) = \sum_{j=0}^{\infty} c_j B^j$, 也就是说,先在(9.2.3)式的形式中寻找最佳设计,而暂不考虑使用反馈信息 ε_t. 最佳设计的标准仍是使相应的输出漂移

$$\varepsilon_t = y_t + y_t^*$$

的方差 $E\varepsilon_t^2$ 达到极小值. 这种馈给系统称为顺馈系统. 我们将证

明，在没有不可量测噪声输入的条件下，最佳顺馈控制就是上一节所要寻求的一般线性最佳控制．下面分两种情况进行讨论．

第一种情况：$b \geqslant f$（一般 $f \geqslant 1$）．由于我们所讨论的系统都是稳定的，所以可取

$$x_t = -\frac{\omega(B)\xi(B)}{\bar{\omega}(B)\zeta(B)} z_t \quad 或 \quad x_t = -\frac{\Omega(B)\xi(B)}{\bar{\Omega}(B)\zeta(B)} B^{b-f} z_t \tag{9.2.4}$$

换句话说，可以设计顺馈系统的传递函数为

$$C(B) = -\frac{\Omega(B)\xi(B)}{\bar{\Omega}(B)\zeta(B)} B^{b-f} \tag{9.2.5}$$

注意 $\Omega(B)\xi(B)$ 和 $\bar{\Omega}(B)\zeta(B)$ 这两个多项式的根都在单位圆外，又 $b-f \geqslant 0$，所以，这个系统是满足稳定性条件的有限阶系统．以它的输出 x_t 为控制量，相应的控制系统的输出是

$$\begin{aligned} y_t^* &= \frac{\bar{\omega}(B)}{\xi(B)} x_t = \frac{\bar{\omega}(B)}{\xi(B)} C(B) z_t = -\frac{\bar{\omega}(B)}{\xi(B)} \frac{\xi(B)\omega(B)}{\zeta(B)\bar{\omega}(B)} z_t \\ &= -\frac{\omega(B)}{\zeta(B)} z_t = -y_t \end{aligned}$$

因此，$\varepsilon_t = y_t + y_t^* = y_t - y_t = 0$．也就是说，在 $b \geqslant f$ 的情况下采用上述顺馈方案时，其输出漂移为零．可见其方差已达到了最小值，即 (9.2.5) 式给出的控制是最佳的．

应当指出，以上只是一种理想化的结果．在实际应用中，即使在 $b \geqslant f$ 的情况下采用了如上的顺馈方案，也还会由于系统模型的各种误差而造成一些残存的输出漂移．因此，为了尽量压缩输出漂移，这种最简单的顺馈控制是不够用的．

第二种情况：$b < f$．这时以 (9.2.5) 式的 $C(B)$ 为传递函数的系统显然不是物理可实现的，因此不能用它作顺馈方案．

为了讨论在这种情况下最佳控制问题的解，我们先假定 z_t 为 ARMA 序列，然后再分析 z_t 为 ARIMA 序列的情形．另外，在讨论过程中，如果不加特别说明，所论及的控制量都是在稳态情况下的控制，即假定 x_t 是由全部历史资料 $z_t, z_{t-1}, \cdots$ 所决定的最

佳线性控制量．这种约定与第七章中平稳预报的理由完全类似．

现在，设 z_t 满足如下的 ARMA 模型：

$$\varphi(B)z_t = \theta(B)a_t \tag{9.2.6}$$

于是相应的工作系统的输出为

$$y_t = \frac{\omega(B)}{\zeta(B)}z_t = \frac{\Omega(B)}{\zeta(B)}B^b z_t = \frac{\Omega(B)}{\zeta(B)}z_{t-b} \tag{9.2.7}$$

而控制系统的输出为

$$y_t^* = \frac{\bar\omega(B)}{\bar\zeta(B)}x_t = \frac{\bar\Omega(B)}{\bar\xi(B)}C(B)B^f z_t = \frac{\bar\Omega(B)}{\bar\zeta(B)}C(B)z_{t-f} \tag{9.2.8}$$

其中 $C(B)$ 是待定的顺馈系统的传递函数．由此可见，y_t^* 只能是 $z_{t-f}, z_{t-f-1}, \cdots$ 的线性组合．依第一章§5和第七章§2可知，为使 $E\varepsilon_t^2 = E(y_t + y_t^*)^2$ 达到极小值，必须且只须

$$y_t^* = -\hat y_t = -E\{y_t|z_{t-f}, z_{t-f-1}, \cdots\} = -E\{y_t|\mathscr{Z}_{t-f}\}$$

其中 $\mathscr{Z}_{t-f}$ 的定义同第一章§5中的类似定义．现在令

$$H(B) = \frac{\Omega(B)}{\zeta(B)} = \sum_{j=0}^{\infty} H_j B^j \tag{9.2.9}$$

那么由(9.2.7)式有

$$\hat y_t = E\{y_t|\mathscr{Z}_{t-f}\} = E\left\{\sum_{j=0}^{\infty} H_j z_{t-b-j}|\mathscr{Z}_{t-f}\right\}$$

$$= \sum_{j=f-b}^{\infty} H_j z_{t-b-j} + \sum_{j=0}^{f-b-1} H_j \hat z_{t-f}(f-b-j)$$

其中 $\hat z_{t-f}(l) = E\{z_{t-f+l}|\mathscr{Z}_{t-f}\}$，是第七章普遍采用的预报记号．由于 z_t 具有(9.2.6)式的模型，按(7.2.4)式可知

$$\hat z_{t-f}(f-b-j) = \pi^{(f-b-j)}(B)z_{t-f}$$

以此式代入前一式得

$$\hat y_t = \left(\sum_{j=f-b}^{\infty} H_j B^{j-(f-b)} + \sum_{j=0}^{f-b-1} H_j \pi^{(f-b-j)}(B)\right) z_{t-f} \tag{9.2.10}$$

注意由(9.2.9)式知

$$Q(B)=\left(\sum_{j=0}^{f-b-1}H_jB^j+\sum_{j=f-b}^{\infty}H_jB^j\right)\zeta(B)$$

$$=\left(\Delta(B)+\sum_{j=f-b}^{\infty}H_jB^j\right)\zeta(B)$$

其中

$$\Delta(B)=\sum_{j=0}^{f-b-1}H_j B^j$$

对比系数可见 $B^{-(f-b)}[Q(B)-\Delta(B)\zeta(B)]=\left(\sum_{j=f-b}^{\infty}H_jB^{j-(f-b)}\right.$

$\left.\cdot\zeta(B)\right)$ 必是 B 的多项式，因而

$$\sum_{j=f-b}^{\infty}H_jB^{j-(f-b)}=B^{-(f-b)}[Q(B)-\Delta(B)\zeta(B)]/\zeta(B)$$

是 B 的有理分式，且分母 $\zeta(B)$ 的根都在单位圆外. 根据类似的推理，由 (7.2.3) 式用数学归纳法，可以证明 $\pi^{(l)}(B)$ 也是这种类型的有理分式. 因此 (9.2.10) 式一定可以表成如下形式:

$$\hat{y}_t=\frac{L_1(B)}{L_2(B)}z_{t-f} \tag{9.2.11}$$

其中 $L_1(B)$ 和 $L_2(B)$ 是 B 的多项式，而且 $L_2(B)$ 的根都在单位圆外. 现在，在 (9.2.8) 式中置 $y_t^*=-\hat{y}_t$，并以 (9.2.11) 式代入，就得到等式

$$-\frac{L_1(B)}{L_2(B)}z_{t-f}=\frac{\bar{Q}(B)}{\bar{\zeta}(B)}C(B)z_{t-f}$$

由此即解出最佳顺馈系统的传递函数为

$$C(B)=-\frac{\bar{\zeta}(B)L_1(B)}{\bar{Q}(B)L_2(B)}\quad 或\quad x_t=-\frac{\bar{\zeta}(B)L_1(B)}{\bar{Q}(B)L_2(B)}z_t \tag{9.2.12}$$

这是一个有限阶的稳定系统，以它的输出 x_t 为控制量，相应控制系统的输出即为 $y_t^*=-\hat{y}_t$，因而使输出漂移 $\varepsilon_t=y_t+y_t^*=y_t-\hat{y}_t$ 的方差达到了极小值. 至于 (9.2.11) 式中 $L_1(B)$ 和 $L_2(B)$

的具体求法，将在后边举例说明，求它们的通式是很烦琐的，我们不在此叙述了.

下面给出在上述顺馈控制下漂移 ε_t 的方差表达式. 由(9.2.6),(9.2.7)和(9.2.9)式

$$
\begin{aligned}
y_t &= H(B)B^b z_t = H(B)\varphi^{-1}(B)\theta(B)B^b a_t \\
&= \delta(B)B^b a_t = \sum_{j=f-b}^{\infty} \delta_j a_{t-b-j} + \sum_{j=0}^{f-b-1} \delta_j a_{t-b-j}
\end{aligned}
$$

再利用最小方差估计的性质，可知

$$
\hat{a}_{t-b-j} = E\{a_{t-b-j} \mid \mathscr{X}_{t-f}\} = \begin{cases} a_{t-b-j}, & j \geqslant f-b \\ 0, & j < f-b \end{cases}
$$

于是

$$
\hat{y}_t = \sum_{j=f-b}^{\infty} \delta_j a_{t-b-j}
$$

由此得

$$
\varepsilon_t = y_t + y_t^* = y_t - \hat{y}_t = \sum_{j=0}^{f-b-1} \delta_j a_{t-b-j} \tag{9.2.13}
$$

$$
E\varepsilon_t^2 = \sum_{j=0}^{f-b-1} \delta_j^2 \sigma_a^2
$$

可见当原系统不加控制时(即取 $x_t = 0$)，其输出漂移的方差为 $\sum_{j=0}^{\infty} \delta_j^2 \sigma_a^2$；而加了最佳顺馈控制以后的剩余漂移的方差比原来减少了 $\sum_{j=f-b}^{\infty} \delta_j^2 \sigma_a^2$.

我们还要说明，上面得出的顺馈方案(9.2.12)式，就是§9.1所要求的一般最佳馈给系统. 也就是说，再根据输出的剩余漂移量 $\varepsilon_t, \varepsilon_{t-1}, \cdots$，无论通过什么样的线性系统构成新的反馈控制分量 $\tilde{x}_t$，加在原来的顺馈控制量 x_t 上，也不可能再进一步降低输出 ε_t 的方差了. 为说明此事实，我们不妨假定相应于 $\tilde{x}_t$ 为输入的控制系统的输出为 $\tilde{y}_t^*$，由于所有系统都是线性的，所以与 $x_t + \tilde{x}_t$ 相

应的控制系统输出为 $y_t^* + \bar{y}_t^*$，因而新的输出漂移量为 $\varepsilon_t^* - \varepsilon_t + \bar{y}_t^*$. 为了使它的方差达到极小值，按照上面类似的论述，应当要求

$$\bar{y}_t^* = -E\{\varepsilon_t | \varepsilon_{t-f}, \varepsilon_{t-f-1}, \cdots\}$$

但是，由于(9.2.13)式可知，此时 ε_t 为 $(f-b-1)$ 阶 MA 序列，故 ε_t 与 ε_{t-f}, ε_{t-f-1}, $\cdots$ 等变量独立，所以

$$\bar{y}_t^* = -E\{\varepsilon_t | \varepsilon_{t-f}, \varepsilon_{t-f-1}, \cdots\} = 0$$

这就说明了 $\bar{x}_t = 0$. 也就等于说，在没有不可量测噪声时，上述顺馈控制就是一种最佳漂移控制.

前面的公式推导中都涉及到输入输出量的无穷多的历史值，但是实际使用的系统总是从有穷时刻开始(一般从 $t=1$ 开始)的，在 $t=0, -1, -2, \cdots$ 等时刻，输入输出量并不知道. 不过，由于所有系统都假定是稳定的和可逆的，所以当系统工作较长时间以后，较早的输入量的作用会逐渐消失. 在这样的条件下，我们不妨认为系统是从无限长历史的零状态开始工作的，即假定 $x_t = z_t = N_t = 0 (t \leqslant 0)$. 这时，上述顺馈控制是渐近最佳的. 当然也可以从 $t=1$ 时刻开始，并考虑其初始值的取法不同而设计出严格的最佳控制方案，那种馈给系统一般不是定常的有限阶系统，而是更复杂的线性系统. 但是当系统工作较长时间以后，两种控制方法的效果就几乎相同了. 这个道理和第七章中序列的平稳预报与适时预报渐近相等的道理是一样的.

现在，再来讨论 z_t 为 ARIMA 序列的情况，即

$$\varphi(B)\nabla^d z_t = \theta(B)a_t, \qquad t \geqslant d+1$$

按照从前的假定，z_t 的初始状态量 $z_1, z_2, \cdots, z_d$ 是具有有穷方差的正态随机变量，它们与 a_t 相互独立. 在这种情况下，系统(9.2.1)总是从 $t=1$ 时刻起以一定的初始状态开始工作的. 为了叙述方便，我们总假定这些初始状态量为零，这等价于补充假定 $z_t = y_t = 0$ $(t \leqslant 0)$. 于是，由(9.2.1)和(9.2.9)式输出 y_t 可形式地表成

$$
\begin{aligned}
y_t &= \sum_{j=0}^{f-b-1} H_j z_{t-b-j} + \sum_{j=f-b}^{\infty} H_j z_{t-b-j} \\
&= \sum_{j=0}^{f-b-1} H_j z_{t-b-j} + \sum_{j=f-b}^{t-b-1} H_j z_{t-b-j}, \quad t > f \qquad (9.2.14)
\end{aligned}
$$

仿 z_t 为 ARMA 序列时的讨论，$\hat{y}_t$ 只能取为在给定 $z_{t-f}, z_{t-f-1}, \cdots, z_1$ 下的最小方差估计，即

$$
\begin{aligned}
\hat{y}_t &= E\{y_t | z_{t-f}, z_{t-f-1}, \cdots, z_1\} \\
&= \sum_{j=f-b}^{t-b-1} H_j z_{t-b-j} + \sum_{j=0}^{f-b-1} H_j E\{z_{t-b-j} | z_{t-f}, z_{t-f-1}, \cdots, z_1\} \\
&\cong \sum_{j=f-b}^{t-b-1} H_j z_{t-b-j} + \sum_{j=0}^{f-b-1} H_j \hat{z}_{t-f}(f-b-j) \qquad (9.2.15)
\end{aligned}
$$

其中 $\hat{z}_{t-f}(f-b-j) = E\{z_{t-b-j} | z_d, z_{d-1}, \cdots, z_1, \mathscr{W}_{t-f}\}$ 是求和序列 z_t 的平稳预报值，其定义见第七章.

例如，若 $d=1$，利用 (7.2.14) 式可知，当 $j < f-b$ 时

$$
\begin{aligned}
\hat{z}_{t-f}(f-b-j) &= z_{t-f} + \sum_{l=1}^{f-b-j} \hat{w}_{t-f}(l) \\
&= z_{t-f} + \sum_{l=1}^{f-b-j} \pi^{(l)}(B) w_{t-f} \\
&\approx z_{t-f} + \sum_{l=1}^{f-b-j} \pi^{(l)}(B) \nabla z_{t-f}
\end{aligned}
$$

其中 $\pi^{(l)}(B)$ 由 (7.2.4) 式定义. 注意由假定 $z_t = 0\ (t \leqslant 0)$ 知 $\nabla^d z_t = 0\ (t \leqslant 0)$，所以 $\pi^{(l)}(B)\nabla^d z_{t-f}$ 是有穷和. 又因 $\nabla^d z_t = w_t\ (t \geqslant d+1)$，可见无穷和 $\pi^{(l)}(B) w_{t-f}$ 与 $\pi^{(l)}(B)\nabla^d z_{t-f}$ 渐近相等. 所以在上面最后一式用了渐近等号"≈". 利用此式和 (9.2.15) 式，以及假定 $z_t = 0\ (t \leqslant 0)$ 可得

$$
\begin{aligned}
\hat{y}_t \approx & \sum_{j=f-b}^{\infty} H_j z_{t-b-j} + \sum_{j=0}^{f-b-1} H_j z_{t-f} \\
& + \sum_{j=0}^{f-b-1} H_j \sum_{l=1}^{f-b-j} \pi^{(l)}(B) \nabla z_{t-f}
\end{aligned}
$$

用前面用过的类似推理可以证明，上式是一个有限阶稳定系统，即可以表为

$$\hat{y}_t \approx \frac{L_1(B)}{L_2(B)} z_{t-f} \qquad (9.2.16)$$

这里的 $L_1(B)$ 和 $L_2(B)$ 的性质与 (9.2.11) 式中的相同. 由此可见，若设计顺馈系统为

$$x_t = C(B)z_t = -\frac{\bar{\zeta}(B)L_1(B)}{\bar{\Omega}(B)L_2(B)} z_t \qquad (9.2.17)$$

那么它必是渐近最佳的.

若 $d=2$，利用 (7.2.15) 式，仿照上面的讨论可以知道，形如 (9.2.17) 式的顺馈控制（当然是不同的 $L_1(B)$ 和 $L_2(B)$）仍是渐近最佳的. 对于一般的 d，也同样可以证明类似的结果. 我们不再一一陈述了.

通过以上的讨论和分析使我们知道，在输入噪声全为可量测的情况下，存在着有限阶的、稳定的渐近最佳顺馈系统. 设此时整个系统的输出为 ε_t，则当系统工作时间足够长以后，ε_t 的方差将达到极小值. 至于这种顺馈系统传递函数 $C(B)$ 的具体求解方法，由于过于烦琐，我们并不去作一般性讨论. 下面将通过一个例子以熟悉顺馈控制方法，并了解具体求出 $L_1(B)$ 和 $L_2(B)$ 的方法要领.

例如，考虑生产某种合成树脂的化学反应系统，要求产品的比重为某个预定值. 输入反应器的某种原料的浓度带有一定的随机起伏，但是这种起伏量可以随时测量，称为可量测噪声 z_t. 在生产过程中，希望根据 z_t 的量值，调节反应器的压力 x_t（称为控制输入量），以使输出产品的比重尽可能稳定在预定值附近. 假定这个化学反应器的系统模型已分别被识别确定为

$$\zeta(B) = 1 - \zeta_1 B, \quad \Omega(B) = 1, \quad b = 0$$

$$\bar{\zeta}(B) = 1 - \bar{\zeta}_1 B, \quad \bar{\Omega}(B) = \bar{\omega}_0, \quad f = 1$$

$$\nabla z_t = a_t - \theta_1 a_{t-1} \quad t \geq 1$$

根据本节前面的论述，约定 $z_t = 0\,(t \leqslant 0)$，于是工作系统的输出为

$$y_t = (1-\zeta_1 B)^{-1} z_t = \sum_{j=0}^{\infty} \zeta_1^j B^j z_t = \sum_{j=0}^{\infty} \zeta_1^j z_{t-j} = z_t + \sum_{j=1}^{\infty} \zeta_1^j z_{t-j}$$

由此可得

$$\begin{aligned}\hat{y}_t &= E\{y_t | z_{t-1}, z_{t-2}, \cdots, z_1\} \\ &= E\{z_t | z_{t-1}, z_{t-2}, \cdots, z_1\} + \sum_{j=1}^{\infty} \zeta_1^j z_{t-j} \\ &\approx \hat{z}_{t-1}(1) + \sum_{j=1}^{\infty} \zeta_1^j z_{t-j}.\end{aligned}$$

利用(7.3.13)式代入上式即有

$$\begin{aligned}\hat{y}_t &\approx z_{t-1} - \theta_1 \sum_{j=0}^{t-3} \theta_1^j (z_{t-j-1} - z_{t-j-2}) + \sum_{j=1}^{\infty} \zeta_1^j z_{t-j} \\ &= z_{t-1} - \theta_1 \sum_{j=0}^{\infty} \theta_1^j z_{t-j-1} + \theta_1 \sum_{j=0}^{\infty} \theta_1^j z_{t-j-2} + \sum_{j=1}^{\infty} \zeta_1^j z_{t-j} \\ &= z_{t-1} - \frac{\theta_1}{1-\theta_1 B} z_{t-1} + \frac{\theta_1}{1-\theta_1 B} z_{t-2} + \frac{\zeta_1}{1-\zeta_1 B} z_{t-1} \\ &= \left(1 - \frac{\theta_1}{1-\theta_1 B} + \frac{\theta_1 B}{1-\theta_1 B} + \frac{\zeta_1}{1-\zeta_1 B}\right) z_{t-1} \\ &= \frac{1-\theta_1+\zeta_1-\zeta_1 B}{(1-\theta_1 B)(1-\zeta_1 B)} z_{t-1} = \frac{L_1(B)}{L_2(B)} z_{t-1}\end{aligned}$$

其中 $L_1(B) = (1-\theta_1+\zeta_1-\zeta_1 B)$，$L_2(B) = (1-\theta_1 B)(1-\zeta_1 B)$。由此可知顺馈系统应为(见(9.2.17)式)

$$x_t = -\frac{(1-\xi_1 B)(1-\theta_1+\zeta_1-\zeta_1 B)}{\bar{\omega}_0(1-\theta_1 B)(1-\zeta_1 B)} z_t \qquad (9.2.18)$$

或者写做

$$(1-\theta_1 B)(1-\zeta_1 B) x_t = \frac{-1}{\bar{\omega}_0}(1-\xi_1 B)(1-\theta_1+\zeta_1-\zeta_1 B) z_t$$

由以上推导可以看到，$L_2(B)$ 是稳定的。因此，上述控制是渐近最佳的。

§3. 反 馈 控 制

现在，我们来讨论 § 9.1 中提出的随机漂移最佳控制问题的另一种特殊情况，即输入噪声全是不可量测噪声的情况(见图9.3.1)。在这种情况下，最佳控制问题相当于在已知 N_t 序列模型和控制系统模型 $y_t^* = \dfrac{\Omega(B)}{\zeta(B)} B^f x_t$ 的条件下，寻找最佳馈给系统 $C(B)$，以使输出漂移 $\varepsilon_t = N_t + y_t^*$ 的方差达到极小值．这时，因为没有可量测噪声输入馈给系统，只有整个系统的输出漂移量 ε_t 做为它唯一的输入量，所以我们把这种馈给系统称为反馈系统，而最佳反馈控制就是 §9.1 所要求的最佳控制．下面就来分析 $C(B)$ 的设计方法．

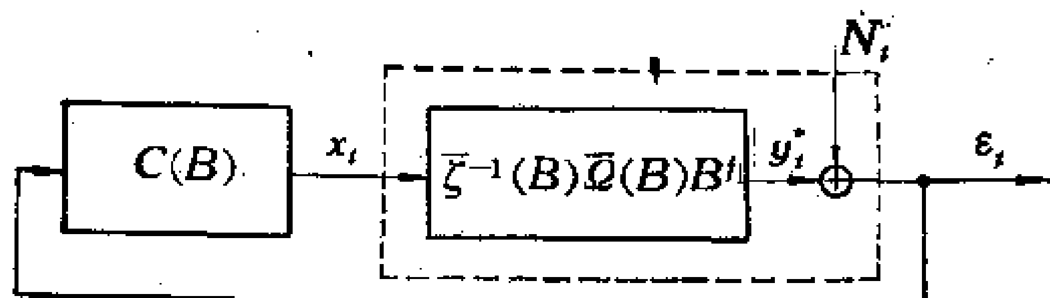


图 9.3.1 反馈控制示意图

我们先考虑 N_t 为 ARMA 序列的情形．与前节类似，由于系统 $C(B)$ 以 ε_t 为其输入量，所以输出量只能是 $\varepsilon_t, \varepsilon_{t-1}, \cdots$ 的线性组合，即 $C(B) = \sum_{j=0}^{\infty} c_j B^j$，其中不能含有 $B^{-k}(k > 0)$ 的项．又由于

$$y_t^* = \frac{\varpi(B)}{\bar{\zeta}(B)} x_t = \frac{\bar{\Omega}(B)}{\bar{\zeta}(B)} B^f x_t = \frac{\bar{\Omega}(B)}{\bar{\zeta}(B)} C(B) \varepsilon_{t-f}$$

所以实际上 y_t^* 只能是 $\varepsilon_{t-f}, \varepsilon_{t-f-1}, \cdots$ 的线性组合．因此，为要使 $\varepsilon_t = N_t + y_t^*$ 的方差达到极小值，必须而且只须

$$y_t^* = -\hat{N}_t^* = -E\{N_t | \varepsilon_{t-f}, \varepsilon_{t-f-1}, \cdots\} \qquad (9.3.1)$$

这个最小方差估计的具体形式一般不易求出．代替它，我们先来考虑如下的预报形式：

$$\hat{N}_t = E\{N_t | N_{t-f}, N_{t-f-1}, \cdots\} = \hat{N}_{t-f}(f) \quad (9.3.2)$$

用 $\varphi(B)N_t = \theta(B)a_t$ 表 N_t 的模型，于是

$$N_t = \varphi^{-1}(B)\theta(B)a_t = \Psi(B)a_t = \sum_{j=0}^{\infty} \psi_j a_{t-j}$$

$$= \sum_{j=f}^{\infty} \psi_j a_{t-j} + \sum_{j=0}^{f-1} \psi_j a_{t-j}$$

$$= \Psi_1(B)B^f a_t + \Psi_2(B)a_t \quad (9.3.3)$$

其中

$$\Psi(B) = \varphi^{-1}(B)\theta(B) = \sum_{j=0}^{\infty} \psi_j B^j$$

$$\Psi_1(B) = \sum_{j=f}^{\infty} \psi_j B^{j-f} = \sum_{j=0}^{\infty} \psi_{j+f} B^j$$

$$\Psi_2(B) = \sum_{j=0}^{f-1} \psi_j B^j = \Psi(B) - \Psi_1(B)B^f$$

由第七章 §2 可知

$$\hat{N}_t = \Psi_1(B)B^f a_t \quad (9.3.4)$$

$$e_t = N_t - \hat{N}_t = \Psi_2(B)a_t \quad (9.3.5)$$

若 $\Psi_2(B)$ 的根都在单位圆外，则由 (9.3.5) 式有

$$a_t = \Psi_2^{-1}(B)e_t$$

以此代入 (9.3.4) 式得

$$\hat{N}_t = \frac{\Psi_1(B)}{\Psi_2(B)} B^f e_t \quad (9.3.6)$$

现在，如果令反馈系统为

$$x_t = C(B)\varepsilon_t = -\frac{\xi(B)\Psi_1(B)}{\Omega(B)\Psi_2(B)} \varepsilon_t \quad (9.3.7)$$

那么，它是一个有限阶的稳定系统. 这时

$$\varepsilon_t = N_t + y_t^* = N_t + \frac{\Omega(B)}{\xi(B)} B^f x_t = N_t - \frac{\Psi_1(B)}{\Psi_2(B)} B^f \varepsilon_t$$

移项得

$$N_t = \left(1 + \frac{\Psi_1(B)}{\Psi_2(B)} B^f\right) \varepsilon_t = \frac{\Psi(B)}{\Psi_2(B)} \varepsilon_t$$

所以图 9.3.1 所示整个系统的输出漂移恰好是

$$\varepsilon_t = \Psi_2(B)\Psi^{-1}(B)N_t = \Psi_2(B)\alpha_t = e_t = N_t - \hat{N}_t \quad (9.3.8)$$

它是一个 $f-1$ 阶的 MA 序列. 由此易见，以 (9.3.7) 式为反馈系统的控制方案一定是最佳的，其理由和用 (9.2.13) 式证明 (9.2.12) 式的最佳性完全类似.

但是这里必须指出，并非任何时候都能保证 (9.3.3) 式中的 $\Psi_2(B)$ 的根都在单位圆外，有时它的根会出现在单位圆内或圆上. 如果发生了这种情况，那么记号 $\Psi_2^{-1}(B)$ 或 $\frac{1}{\Psi_2(B)}$ 就没有确切的数学含义了. 这时我们不能使用 (9.3.7) 式来表达反馈系统，但仍可用一种表面上与它等价的形式表达一个非稳定的系统，即

$$\Omega(B)\Psi_2(B)x_t = -\xi(B)\Psi_1(B)\varepsilon_t, \quad t > t_0 \quad (9.3.9)$$

这个等式表示在任意给定初值 $\varepsilon_t, x_t\ (t \leqslant t_0)$ 以后，输入 ε_t 与输出 x_t 当 $t > t_0$ 时的递推关系. 为方便起见，我们总约定取 $t_0 = 0$, $\varepsilon_t = x_t = 0\ (t \leqslant 0)$；这时，(9.3.9) 式事实上对 $t \leqslant 0$ 也是成立的. 现在，若以 (9.3.9) 式做为图 9.3.1 所示系统的反馈系统，但须记住这时它并不具有稳定性，而且其中的 ε_t 只是从时刻 $t \geqslant 1$ 起才真正等于整个系统的输出漂移量 $N_t + y_t^*$(在此以前，$\varepsilon_t = 0$)，于是由此可得

$$\begin{aligned}
\Psi_2(B)\varepsilon_t &= \sum_{j=0}^{f-1} \phi_j \varepsilon_{t-j} = \sum_{j=0}^{f-1} \phi_j(N_{t-j} + y_{t-j}^*) \\
&= \Psi_2(B)N_t + \Psi_2(B)y_t^* \\
&= \Psi_2(B)N_t + \Psi_2(B)\frac{\Omega(B)}{\xi(B)} B^f x_t \\
&= \Psi_2(B)N_t - \Psi_1(B)B^f\varepsilon_t \qquad t \geqslant f
\end{aligned}$$

移项得

$$\Psi(B)\varepsilon_t = (\Psi_2(B) + \Psi_1(B)B^f)\varepsilon_t = \Psi_2(B)N_t, \quad t \geqslant f \quad (9.3.10)$$

由于 $\Psi(B) = \varphi^{-1}(B)\theta(B)$，当 $t \geqslant p + f$ 时，在上式两边同施

以运算 $\varphi(B)$，则又有

$$\theta(B)\varepsilon_t=\varphi(B)\Psi_2(B)N_t=\Psi_2(B)\theta(B)\alpha_t,\quad t\geqslant p+f \tag{9.3.11}$$

其中 p 为 $\varphi(B)$ 的次数. 注意此式仅当 $t\geqslant p+f$ 才成立，所以不能由两边施行运算 $\theta^{-1}(B)$ 推出 $\varepsilon_t=\Psi_2(B)\alpha_t=N_t-\hat{N}_t$. 但是，若记 $\theta^{-1}(B)=\sum_{j=0}^{\infty}q_jB^j$，则必成立恒等式

$$\xi_t=\theta^{-1}(B)\theta(B)\xi_t=\sum_{j=0}^{\infty}q_j\theta(B)\xi_{t-j}$$

其中 ξ_t 是任意平稳序列. 于是，由 (9.3.11) 式可得

$$\begin{aligned}\varepsilon_t&=\sum_{j=0}^{\infty}q_j\theta(B)\varepsilon_{t-j}=\sum_{j=0}^{t-p-f}q_j\Psi_2(B)\theta(B)\alpha_{t-j}\\&\quad+\sum_{j=t-p-f+1}^{\infty}q_j\theta(B)\varepsilon_{t-j}=\sum_{j=0}^{\infty}q_j\theta(B)\Psi_2(B)\alpha_{t-j}\\&\quad+\sum_{j=t-p-f+1}^{\infty}q_j\theta(B)[\varepsilon_{t-j}-\Psi_2(B)\alpha_{t-j}]\\&=\Psi_2(B)\alpha_t+\sum_{j=t-p-f+1}^{\infty}q_j\theta(B)[\varepsilon_{t-j}-\Psi_2(B)\alpha_{t-j}]\end{aligned}$$

因为 q_j 是以 j 的负指数幂下降的，所以当 $t\to\infty$ 时，

$$E[\varepsilon_t-\Psi_2(B)\alpha_t]^2=E\left\{\sum_{j=t-p-f+1}^{\infty}q_j\theta(B)[\varepsilon_{t-j}-\Psi_2(B)\alpha_{t-j}]\right\}^2\to 0$$

由类似于 (9.3.8) 式的论证可见，即使是 $\Psi_2(B)$ 的根不全在单位圆外的情形，若采用 (9.3.9) 式作为非稳定的反馈系统，它仍然是渐近最佳的.

如果将图 9.3.1 所示带最佳反馈的整个系统看做以 N_t 为输入，ε_t 为输出的系统，那么，先不管 $\Psi_2(B)$ 是什么情况，由 (9.3.8) 和 (9.3.10) 式，这个系统可用下式描述：

$$\Psi(B)\varepsilon_t=\Psi_2(B)N_t$$

或

$$\theta(B)\varepsilon_t = \Psi_2(B)\varphi(B)N_t \qquad (9.3.12)$$

因为 $\theta(B)$ 的根已假定全在单位圆外，所以按照第八章§1关于系统稳定性的定义，(9.3.12) 式是稳定的．然而，尽管整个系统的输入输出具有 (9.3.12) 式这样稳定的传递形式，但它的内部结构却仍可能呈现不稳定性．如图 9.3.1，整个系统是由反馈系统与控制系统串联构成的，只有这两个系统都稳定，它们才能付之使用．其原因在于，实际应用的系统不可避免地会存在各种各样的误差因素，诸如初值的影响，线性化描述的误差，传递回路（或计算）误差等等．这些误差会由于系统的不稳定性而导致无限制的积累，以至发散．以非稳定的反馈系统 (9.3.9) 式为例，如果它含有干扰误差项 η_t，即

$$\bar{Q}(B)\Psi_2(B)x_t = -\bar{\zeta}(B)\Psi_1(B)\varepsilon_t + \eta_t, \quad t \geqslant 1 \quad (9.3.13)$$

η_t 相对于 ε_t 是高阶小量，而且可以近似当做平稳序列．若以 x_t^0 表示反馈系统设计的理论输出值，即 x_t^0 是严格满足 (9.3.9) 式的，那么 (9.3.9) 与 (9.3.13) 两式相减可得

$$\bar{Q}(B)\Psi_2(B)(x_t - x_t^0) = \eta_t, \quad t \geqslant 1 \qquad (9.3.14)$$

由此可见，实际控制量与其设计值的误差 $\tilde{x}_t = x_t - x_t^0$ 是一个非平稳序列，因此它的方差一般说来是发散的，即随着系统工作时间的延长，它的方差会越来越大．但另一方面，按照前面的论证，上述的反馈控制方案又是渐近最佳的，因此工作时间越长，才越能显示其优效性质．这就导致了应用上的矛盾！在顺馈控制中并无这种现象发生．总之，在使用反馈控制时，虽然无论 $\Psi_2(B)$ 是否满足稳定性条件，其反馈系统设计基本相同．但是在 $\Psi_2(B)$ 不满足稳定性条件时，反馈控制的应用效果就要受到工作时间的限制，甚至达到不能采用的程度．

例如设有一生产人造纤维的聚合反应器，要求输出半成品的粘度为某值．由于反应器进料成分有一定的波动，导致了粘度的漂移．这种波动不能及时化验，这相当于一种不可量测的输入噪声．经试验分析后，已经给出了在非控制情况下输出半成品的粘

度漂移量 N_t 的随机模型．现在要设计一个调节催化剂加入量的馈给系统，以使控制后的漂移量 ε_t 的方差尽量小；而催化剂对于输出漂移系统(即控制系统)的关系经识别确定为 $y_t^* = \varpi_0 x_{t-2}$，即

$$\zeta(B) = 1, \qquad \Omega(B) = \varpi_0, \qquad f = 2$$

N_t 的模型为

$$N_t = (1 - \theta_1 B - \theta_2 B^2)\alpha_t$$

此时，依 (9.3.3) 式

$$N_t = \Psi(B)\alpha_t = \Psi_1(B)\alpha_{t-1} + \Psi_2(B)\alpha_t$$
$$= -\theta_2\alpha_{t-2} + (1 - \theta_1 B)\alpha_t$$

因此由 (9.3.9) 式，最佳反馈控制系统为

$$\varpi_0(1 - \theta_1 B)x_t = \theta_2\varepsilon_t, \quad t \geqslant 1 \tag{9.3.15}$$

而且控制后的漂移量近似为

$$N_t - \hat{N}_t = \Psi_2(B)\alpha_t = \alpha_t - \theta_1\alpha_{t-1}$$

其方差为 $(1 + \theta_1^2)\sigma_a^2$，比未加控制时的方差 $EN_t^2 = (1 + \theta_1^2 + \theta_2^2)\sigma_a^2$ 减少了 $\theta_2^2\sigma_a^2$.

需要指出的是，当 $|\theta_1| < 1$ 时，反馈系统 (9.3.15) 式是稳定的，这时，(9.3.7) 式反馈系统的传递函数可用

$$C(B) = -\frac{\zeta(B)\Psi_1(B)}{\Omega(B)\Psi_2(B)} = \frac{\theta_2}{\varpi_0}\frac{1}{(1 - \theta_1 B)}$$

表示之；当 $|\theta_1| \geqslant 1$ 时，则是不稳定的．比如当 $\theta_1 = 1$，$\theta_2 = -0.24$，$\varpi_0 = 1$ 时 (9.3.15) 式为

$$(1 - B)x_t = -0.24\varepsilon_t, \qquad t \geqslant 1$$

这是一个一阶求和系统，因此，其实际输出控制量 x_t 与其设计值 x_t^0 之差满足(见(9.3.14)式)

$$(1 - B)\tilde{x}_t = \eta_t, \qquad t \geqslant 1$$

由此解得 $\tilde{x}_t = \sum_{j=1}^{t}\eta_j$．若 η_t 为一平稳白噪声，则 $E\tilde{x}_t^2 = tE\eta_j^2$；若 η_t 含有固定的常数误差项 η，则 x_t 也将含有非随机误差项 $t\eta$，它们都会随着 t 的增长而无限变大．由此可见不稳定系统会给实际应用带来多大的麻烦！

最后，我们提供一种方法来弥补上述反馈控制中的不足之处，供读者参考．当 $\Psi_2(B)$ 不满足稳定性条件时，可以采用复现控制系统的输出量 y_t^* 的办法，来设计如图 9.3.2 所示的反馈系统．这时，系统 $C(B)$ 的输入是 N_t，因此

$$y_t^* = \frac{\bar{\omega}(B)}{\bar{\zeta}(B)} x_t = \frac{\bar{Q}(B)}{\bar{\zeta}(B)} C(B) N_{t-f}$$

为了使 $\varepsilon_t = N_t + y_t^*$ 达极小值，只须取

$$y_t^* = -\hat{N}_t = -E\{N_t | N_{t-f}, N_{t-f-1}, \cdots\} = -\pi^{(f)}(B) N_{t-f}$$

即取

$$C(B) = -\frac{\bar{\zeta}(B)}{\bar{Q}(B)} \pi^{(f)}(B)$$

我们在 (9.2.11) 式已经指出过，$\pi^{(f)}(B)$ 是稳定的，我们又总假

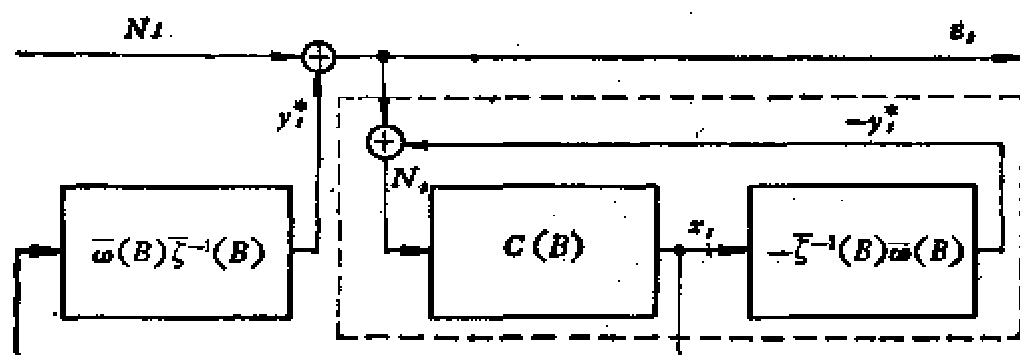


图 9.3.2　复现 y_t^* 的反馈控制示意图

定 $\bar{Q}(B)$ 的根在单位圆外，所以 $C(B)$ 是稳定的，而且控制后的输出漂移 ε_t 仍如 (9.3.8) 式所示．容易验证，这种方法对 N_t 为 ARIMA 序列时也同样适用．在此必须指出，在图 9.3.2 中虚线外的系统 $\frac{\bar{\omega}(B)}{\bar{\zeta}(B)}$，只是真实操作过程（如化学反应过程）的传递函数的近似描述，而虚线内的复现系统 $-\frac{\bar{\omega}(B)}{\bar{\zeta}(B)}$，则完全是按这个近似模型设计的．因此实际上后者的输出不可能完全复现真实操作过程的输出．容易想见，$\frac{\bar{\omega}(B)}{\bar{\zeta}(B)}$ 对真实操作过程的系统拟合得越好，图 9.3.2 的控制方法的效果也就越好．当然，不

只对这种控制如此，对于图 9.2.1 和图 9.3.1 的控制而言也是这样；不能期望基于对真实工作过程的误差很大的系统描述，能建立很好的控制方案.

§4. 顺馈-反馈联合控制

在前两节里，我们讨论了随机漂移控制问题的两种特殊情况。现在开始讨论它的一般情况的解决办法. 在一般情况下，系统的输入噪声既有可量测的，又有不可量测的，这时只用反馈或只用顺馈方法都达不到最佳控制的要求. 按照本章 §1 关于随机漂移控制问题的概述，我们假定图 9.4.1 中的控制系统与工作系统的传递函数都是已知的，此外，可量测输入噪声 z_t 以及不可量测噪声所产生的输出量 N_t(见图 9.1.1 所示)的随机模型也都已获得，又假定它们是相互独立的序列. 这时最佳馈给系统是以 z_t 和 ε_t 为双输入的线性系统. 如图 9.4.1 所示，我们把这种馈给系统分成两个并联的单输入单输出的有限阶系统.

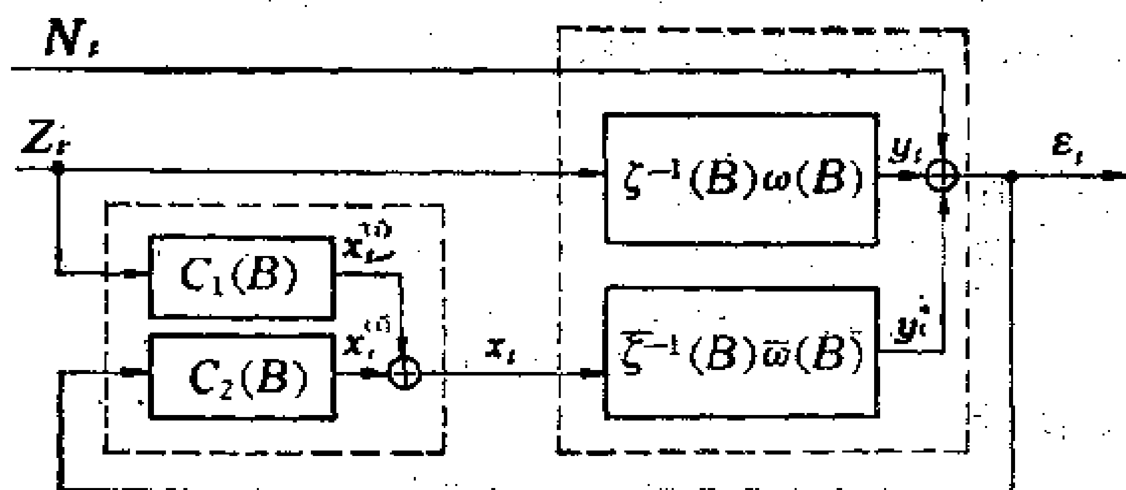


图 9.4.1 顺馈-反馈联合控制示意图

然后分别根据本章 §2 和 §3 中的原理，设计图 9.4.1 中的馈给系统的顺馈部分 $C_1(B)$ 和反馈部分 $C_2(B)$，称这种控制为顺馈-反馈联合控制. 首先在未加反馈控制的条件下来设计顺馈部分. 这时由于 z_t 与 N_t 是相互独立的，因此 y_t, y_t^* 与 N_t 也独立，于是，为了使 $y_t^* + y_t + N_t$ 的方差达到极小值，只须考虑使 $y_t^* + y_t$ 达

到极小值，于是可按照 §9.2 中的方法设计 $C_1(B)$. 设它的有限阶表示为（见（9.2.12）或（9.2.17）

$$C_1(B) = -\frac{\bar{\zeta}(B)L_1(B)}{\bar{\Omega}(B)L_2(B)}$$

在施加这个顺馈控制后，系统的输出漂移量为

$$n_t = N_t + y_t - \hat{y}_t$$

又由（9.2.13）式知

$$y_t - \hat{y}_t = \sum_{j=0}^{t-b-1} \delta_j a_{t-b-j}$$

由于 N_t 与 $y_t - \hat{y}_t$ 相互独立，按照第六章 §1 可知，合成序列 $n_t = N_t + y_t - \hat{y}_t$ 也是 ARMA 序列，其模型由第六章 §1 的方法便可获得. 于是将 n_t 视为 § 9.3 中的 N_t，并按那里的原则设计出反馈部分 $C_2(B)$ 来，就是我们所需要的. 以上是在一般情况下最佳随机漂移控制的一种设计方案.

例如，继续考虑本章 § 2 的例子，假定在前述生产合成树脂的过程中，同时存在一项不可量测的噪声输入，它对输出漂移的影响量为 N_t，其模型被识别为 AR(1, 0) 型：

$$N_t - \varphi_1 N_{t-1} = \alpha_t$$

其它条件仍如前述. 现在要求设计一个顺馈-反馈联合控制方案. 按照本节所介绍的设计方法，顺馈部分仍然如旧，即为(9.2.18)式：

$$x_t^{(1)} = -\frac{(1-\zeta_1 B)(1-\theta_1+\zeta_1-\zeta_1 B)}{\bar{\omega}_0(1-\theta_1 B)(1-\zeta_1 B)} z_t$$

或者表为

$$C_1(B) = -\frac{(1-\zeta_1 B)(1-\theta_1+\zeta_1-\zeta_1 B)}{\bar{\omega}_0(1-\theta_1 B)(1-\zeta_1 B)}$$

在顺馈控制下系统的输出漂移量为

$$n_t = N_t + y_t - \hat{y}_t$$

由于系统是从 $t = 1$ 时刻开始工作的，所以

$$y_t = z_t + \sum_{j=1}^{t-1} \zeta^j z_{t-j} = z_t + \sum_{j=1}^{\infty} \zeta^j z_{t-j}$$

其中约定 $z_j=0\,(j\leqslant 0)$. 又有

$$\hat{y}_t=\hat{z}_{t-1}(1)+\sum_{j=1}^{t-1}\zeta_j z_{t-j}=\hat{z}_{t-1}(1)+\sum_{j=1}^{\infty}\zeta_j z_{t-j}$$

注意 z_t 已被识别为 ARIMA (0, 1, 1) 序列:

$$\nabla z_t=a_t-\theta_1 a_{t-1}$$

从而

$$y_t-\hat{y}_t=z_t-\hat{z}_{t-1}(1)=a_t$$

又由于 N_t 是 AR(1, 0) 序列, 且与 a_t 独立, 根据(6.2.12)和(6.2.5)等式可知 $n_t=N_t+a_t$ 是 ARMA(1, 1)序列,其模型为

$$(1-\bar{\varphi}_1 B)n_t=(1-\bar{\theta}_1 B)b_t$$

其中 b_t 表示与 n_t 相应的模型噪声序列, 而参数 $\bar{\theta}_1$ 和 σ_b^2 由以下方程式解出

$$(1+\bar{\theta}_1^2)\sigma_b^2=\sigma_a^2+(1+\bar{\varphi}_1^2)\sigma_a^2$$

$$-\bar{\theta}_1\sigma_b^2=-\bar{\varphi}_1\sigma_a^2$$

其中 $\bar{\varphi}_1\,\sigma_\varepsilon^2$, σ_a^2 都是已知参数. 现在将 n_t 代替(9.3.3)式中的 N_t, 写出其传递形式

$$n_t=(1-\bar{\varphi}_1 B)^{-1}(1-\bar{\theta}_1 B)\alpha_t=\sum_{j=0}^{\infty}\psi_j\alpha_{t-j}$$

$$=\alpha_t+\sum_{j=1}^{\infty}\psi_j\alpha_{t-j}=\alpha_t+\sum_{j=1}^{\infty}\bar{\varphi}_1^{j-1}(\bar{\varphi}_1-\bar{\theta}_1)\alpha_{t-j}$$

$$=\Psi_2(B)\alpha_t+\Psi_1(B)B\alpha_t$$

其中

$$\Psi_2(B)=1,\quad \Psi_1(B)=\sum_{j=1}^{\infty}\bar{\varphi}_1^{j-1}(\bar{\varphi}_1-\bar{\theta}_1)B^{j-1}$$

$$=(\bar{\varphi}_1-\bar{\theta}_1)\sum_{j=0}^{\infty}\bar{\varphi}_1^j B^j=\frac{\bar{\varphi}_1-\bar{\theta}_1}{1-\bar{\varphi}_1 B}$$

因此,根据(9.3.7)式,反馈部分的传递函数应为

$$C_2(B)=-\frac{(1-\xi B)}{\bar{\omega}_0}\frac{\Psi_1(B)}{\Psi_2(B)}=-\frac{(\bar{\varphi}_1-\bar{\theta}_1)}{\bar{\omega}_0}\frac{(1-\xi B)}{(1-\bar{\varphi}_1 B)}$$

在此例中, 由于 $\Psi_2(B)=1$, 所以 $C_2(B)$ 是稳定系统, 而

$C_1(B)$ 总是稳定的. 至于在一般情况下，关于反馈部分的稳定性的讨论和本章 § 3 完全相似，这里不再重复了. 通过此例我们还可以看到，由于可量测噪声 z_t 是非平稳的 ARIMA (0, 1, 1) 序列，所以在不施加任何控制手段时（即恒取 $x_t = 0$)，输出随机漂移 $N_t + y_t$ 也是非平稳序列，因而它的方差会越来越大，以至达到实际应用所不能允许的地步. 但是在加了顺馈控制后，其输出漂移 $n_t = N_t + y_t - \hat{y}_t$ 就变为平稳 ARMA 序列了. 如果再采取顺馈-反馈联合控制，则其输出漂移的方差又进一步变小.

我们在此顺便指出，如果实际系统有 m 个独立的可量测输入噪声，而且各自对输出漂移的"贡献"表现为 m 个单输入单输出系统的并联叠加(见图9.4.2)，那么，本章的各种控制方法容易推广到这种多维简单并联的系统的情况. 其结果相当于独立设计 m 个顺馈系统，最后再并联一个反馈系统. 不难想象，这种推广也同样适用于存在多道控制系统的情况.

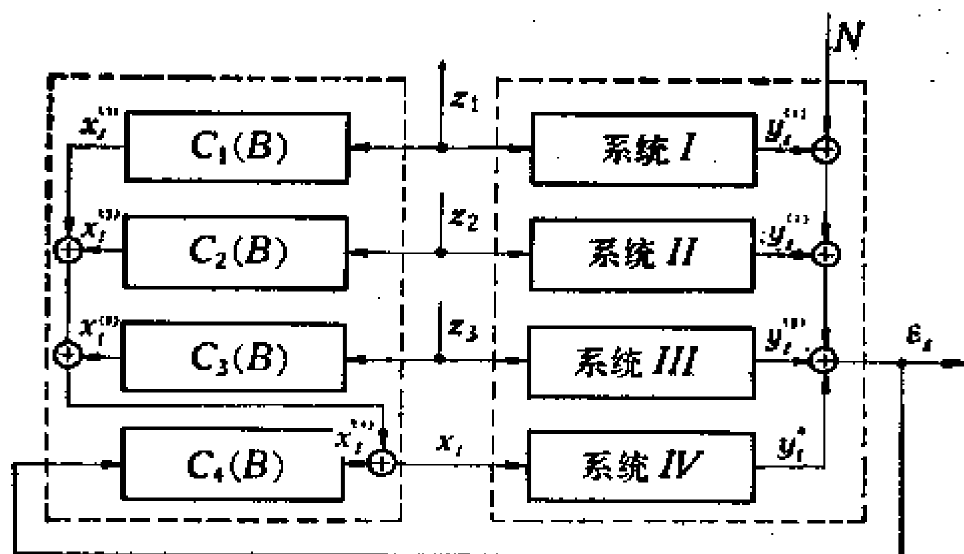


图 9.4.2 三维顺馈-反馈控制系统示意图

第十章 极大熵谱估计与AR模型

§1. 以极大熵为准则的谱估计

一、谱估计

在工程技术和某些科学领域中，存在着大量谱估计的问题. 为了描述一个实在的随机序列，在时间域内，我们可以借助于相关性建立起 ARMA 这类形式的模型. 我们也可从频率特性方面去描述，因此谱估计理论在时间序列的统计理论中是另一个基本的方面.

在第三章 §3 中我们已看到，平稳序列的相关函数和谱函数存在着互相转换的关系式（参见（1.3.7）～（1.3.8）式）. 但是，它们的估计量就不一定满足这些关系，因为估计值是随机变量. 对于估计量，我们应当追求它有好的渐近性质，尤其在大样本的情况下更是如此. 例如用

$$\hat{\gamma}_k = \frac{1}{n}\sum_{t=1}^{n-k} x_t x_{t+k}, \quad \hat{\gamma}_{-k} = \hat{\gamma}_k \tag{10.1.1}$$

估计平稳序列 x_t 的协方差函数 γ_k，在适当条件下，$\hat{\gamma}_k$ 具有渐近无偏性、相容性等性质. 那么根据谱密度 $g(\lambda)$ 和 γ_k 的关系（见（1.3.9）式），$g(\lambda)$ 的估计量可取做

$$\hat{g}_n(\lambda) = \sum_{k=-n+1}^{n-1} \hat{\gamma}_k e^{-i2\pi k\lambda} \tag{10.1.2}$$

若 $\hat{\gamma}_k$ 如（10.1.1）式所定义，则

$$\hat{g}_n(\lambda) = \left(\frac{1}{n}\sum_{t=1}^{n} x_t^2\right) + 2\sum_{k=1}^{n-1}\left(\frac{1}{n}\sum_{t=1}^{n-k} x_t x_{t+k}\right)\cos 2\pi\lambda k$$
$$= I_n(\lambda) \tag{10.1.3}$$

从第五章 §3 中看出（见（5.3.4）式），$\hat{g}_n(\lambda)$ 是 $g(\lambda)$ 的渐近无偏估计. 但是当 $n\to\infty$ 时，$E\hat{g}_n^2(\lambda)$ 的均方根值却不能收敛于 0（没

有相容性)，甚至比 $g(\lambda)$ 本身的值还可能大，因此用 (10.1.2) 式定义的 $\hat{g}(\lambda)$ 来估计 $g(\lambda)$ 并不理想.

长时间以来在努力寻求较为理想的谱估计的过程中，人们注意到，虽然 $\hat{r}_0, \hat{r}_1, \cdots$ 的误差方差随着足标的增加而变大，但用 (10.1.2) 式估计 $g(\lambda)$ 时未考虑到这一事实，因此造成了 $\hat{g}(\lambda)$ 的方差不趋于 0. 从而人们自然会想到，在 (10.1.2) 式中给每个 $\hat{r}_k$ 以适当的"权系数" $w_k^{(n)}$，使得不同的 $\hat{r}_k$ 对 $\hat{g}(\lambda)$ 的贡献大小有所区别，即

$$\hat{g}_n^*(\lambda) = \sum_{k=-n+1}^{n-1} w_k^{(n)} \hat{r}_k e^{-i2\pi k\lambda} \tag{10.1.4}$$

当然，$|w_k^{(n)}|$ 应随着 $|k|$ 的增长越来越小 (甚至可以等于零). (10.1.4) 式的加权估计 $\hat{g}_n^*(\lambda)$ 实际上是对 $g(\lambda)$ 的某种滑动平均值的近似估计. 这可从下面的变换中看出. 先令

$$K_n(\omega) = \sum_{k=-n+1}^{n-1} w_k^{(n)} e^{-i2\pi k\omega}$$

由 (10.1.4) 式，有

$$\hat{g}_n^*(\omega) = \int_{-\frac{1}{2}}^{\frac{1}{2}} K_n(\lambda)\hat{g}_n(\omega-\lambda)d\lambda$$

$$\begin{aligned} E\hat{g}_n^*(\omega) &= \int_{-\frac{1}{2}}^{\frac{1}{2}} K_n(\lambda)g(\omega-\lambda)d\lambda \\ &\quad - \int_{-\frac{1}{2}}^{\frac{1}{2}} K_n(\lambda)[E\hat{g}_n(\omega-\lambda)-g(\omega-\lambda)]d\lambda \end{aligned}$$

知 $\hat{g}_n^*(\omega)$ 是对

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} K_n(\lambda)g(\omega-\lambda)d\lambda$$

的近似估计(代替对 $g(\omega)$ 本身的估计). 若将 $\hat{g}_n^*(\omega)$ 作为 $g(\omega)$ 的估计，则有偏差 (设 $w_0^{(n)}=1$).

$$b_n(\omega) \cong \int_{-\frac{1}{2}}^{\frac{1}{2}} K_n(\lambda)g(\omega-\lambda)d\lambda - \int_{-\frac{1}{2}}^{\frac{1}{2}} K_n(\lambda)g(\omega)d\lambda \tag{10.1.5}$$

虽然 $b_n(\omega)$ 一般不等于零，但如果适当选择 $w_k^{(n)}$，则可使 $\mathrm{var}\hat{g}_n^*(\omega)$ 趋于零，其估计误差的均方差为

$$E[\hat{g}_n^*(\omega)-g(\omega)]^2=\mathrm{var}\hat{g}_n^*(\omega)+b_n^2(\omega) \quad (10.1.6)$$

可以比 $E[\hat{g}_n(\omega)-g(\omega)]^2$ 小得多．这种方法在文献中称为谱窗估计，$K_n(\omega)$ 称为窗函数．谱窗估计的技巧在于如何选取 $w_k^{(n)}$（亦即 $K_n(\omega)$），使 $E[\hat{g}_n^*(\omega)-g(\omega)]^2$ 尽快趋于零，于是产生了各种各样的窗．人们可以根据经验选取某种形状某种宽度的窗（即适当地权衡 var $\hat{g}_n^*(\omega)$ 和 $b_n(\omega)$ 的大小）给出合适的谱估计．读者可参看 [11] [12].

用谱窗方法常常可以获得谱的相容性估计，但是，它有以下缺点：

(1) 它是一种非参数化的估计方法，须对逐点 λ 给出 $g(\lambda)$ 的估计值 $\hat{g}_n^*(\lambda)$，因而只适合于计算机的直接运算，不便用来作分析推导或建立数学模型.

(2) 一般只在大样本的情况下能算出偏差 $b_n(\lambda)$ 和方差 var $\hat{g}_n^*(\lambda)$的近似值，在小样本情况下，不便评价一个谱窗的优劣.

(3) 谱窗估计对隐含在随机序列中的谐波分量（如 $A\sin\omega t$ 分量）不太敏感，也就是说不容易根据 $\hat{g}_n^*(\lambda)$ 的峰值点分辨出隐含的谐波．特别当采样时间 T 小于隐含谐波的周期长度 T_0（即若以 Δ 表示采样间隔，n 表示样本个数，T_0 表示隐含谐波的周期，$T=n\Delta<T_0$），还容易产生峰值移动，也就是说在 $\lambda_0=\frac{1}{n}\frac{T}{T_0}$ 处 $\hat{g}_n^*(\lambda)$ 应有的高峰实际上却移到别处去了（详见[13]）.

除了谱窗估计方法外，人们还用其他方法来估计谱．前面几章建立的 ARMA 模型参数的估计方法，实际上也是一种谱估计的方法．ARMA 序列的谱密度是一种有理函数的形式．在时间域内用 ARMA 模型 $\varphi(B)x_t=\theta(B)a_t$ 来拟合一个平稳序列，相当于在频率域内用有理函数（见(1.3.18)式）

$$g(\lambda)=\sigma_a^2\left|\frac{\theta(e^{i2\pi\lambda})}{\varphi(e^{i2\pi\lambda})}\right|^2 \quad (10.1.7)$$

来估计序列的谱密度．这种估计的特点是使谱参数化了.

我们将证明，在一定的条件下，使谱的熵达极大的谱估计准则

是求一个高阶的 AR 的模型，用它的谱来作为被考察的实际序列的谱估计.

由于它对隐含的谐波分量有高度敏感性. 因此极大熵谱估计方法已越来越多地在工程和很多科研领域中被采用. 但是，在使用熵谱分析时，如何选取 AR 的阶数是个困难的问题.

二、谱熵

一般说来，如果一个随机序列相关性越强，那么它的随机性越弱；反之相关性越弱，随机性越强. 正态白噪声是一种“最随机”的序列，因为它在任意两个不同时刻上的取值毫不相关，即使知道了某些时刻序列的取值，对于估计其他时刻序列的取值毫无参考意义.

考察所有均值为零方差为 γ_0 的平稳序列 x_t. 正如第七章 §4 中所定义的，我们称

$$\varepsilon_{k+1} = x_{k+1} - \hat{E}(x_{k+1}|x_1, \cdots x_k) \qquad (\varepsilon_1 = x_1)$$

为 x_{k+1} 带来的新息. 随机性越大，新息的方差 σ_k^2 也越大 ($\sigma_k^2 \leqslant \gamma_0$). 因此，我们用

$$Q = \sum_{k=1}^{n} \sigma_k^2$$

来衡量长度为 n 的样本所包含的新息量的多少. 显然，x_t 的随机性越大，Q 也越大. 这就是说，随机性越强的序列包含着越多的新息. 白噪声使 Q 达极值 $Q = n\gamma_0$，所以对于同样方差的平稳序列，白噪声的 n 个样本比其他序列的 n 个样本包含了最多的信息量.

设 $g(\lambda)$ 是一个序列的谱密度，$g(\lambda) > 0 \left(-\frac{1}{2} \leqslant \lambda \leqslant \frac{1}{2}\right)$，我们称

$$H = \int_{-\frac{1}{2}}^{\frac{1}{2}} \log g(\lambda) d\lambda \tag{10.1.8}$$

为此序列的谱熵.

现考虑所有具有正则的谱密度 $g(\lambda)$ (即$\int \log g(\lambda)d\lambda > -\infty$) 的平稳序列，在 $g(\lambda)$ 满足约束条件

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} g(\lambda)d\lambda = \gamma_0 \tag{10.1.9}$$

时(即方差为给定值),怎样的 $g(\lambda)$ 使 (10.1.8) 式中的 H 达极大. H 可写成 $H(g)$，暂不考虑 g 的非负性，熟悉泛函分析的读者可知，H 是定义于$\left[-\frac{1}{2},\frac{1}{2}\right]$的连续函数空间(记为 $\mathscr{C}^0$) 上的泛函.

现在讨论 $H(g)$ 的极值. 设

$$J(g) = \int_{-\frac{1}{2}}^{\frac{1}{2}} \log g(\lambda)d\lambda - \zeta\left[\int_{-\frac{1}{2}}^{\frac{1}{2}} g(\lambda)d\lambda\right]$$

$$= \int_{-\frac{1}{2}}^{\frac{1}{2}} [\log g(\lambda) - \zeta g(\lambda)]d\lambda \tag{10.1.10}$$

其中 ζ 是 Lagrange 乘子. 由变分法知道,对泛函 $J(g)$ 的变分

$$\delta J(g) = \left[\frac{\partial}{\partial \alpha} J(g+\alpha\delta g)\right]_{\alpha=0}$$

其中 δg 表示 g 在空间 $\mathscr{C}^0$ 中的任一"增量". 由此可知(10.1.10) 式所给出的泛函的变分为

$$\delta J = \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{\partial}{\partial \alpha}[\log(g + \alpha\delta g) - \zeta(g + \alpha\delta g)]_{\alpha=0} d\lambda$$

$$= \int_{-\frac{1}{2}}^{\frac{1}{2}} \left[\frac{1}{g(\lambda)} - \zeta\right] \delta g(\lambda)d\lambda \tag{10.1.11}$$

由于达极值的条件是 $\delta J = 0$ 及 $\delta g(\lambda)$ 之任意性，知 $\frac{1}{g(\lambda)} - \zeta = 0$,即

$$g(\lambda) = \frac{1}{\zeta} \tag{10.1.12}$$

使 $J(g)$ 达极值,即 $g(\lambda)$ 必须是白噪声的谱密度. 将 (10.1.12) 式代回到 (10.1.9) 式,最后可得

$$g(\lambda) = \gamma_0 \tag{10.1.13}$$

其非负性也自然满足了.

由此可见，随机性最强的白噪声序列其谱熵也最大. 也就是说，随机的大小和新息量的大小以及谱熵的大小，这三者的变化趋势是一样的. 谱熵是从谱的角度给出了衡量这一趋势的定量表达式. 熟悉平稳过程中 Wold 分解的读者可知，(10.1.8) 式所给出的"谱熵"正好反映了一步平稳预报误差的方差（见 [15]），其方差为 $e^{\int_{-\frac{1}{2}}^{\frac{1}{2}}\log g(\lambda)d\lambda}$.

三、以极大熵为准则的谱估计

设均值为零的平稳序列 x_t 的谱在 $\left[-\frac{1}{2}, \frac{1}{2}\right]$ 上非负、连续. 如果已知 x_t 的自协方函数的前 $m+1$ 个值，

$$\gamma_{-k}=\gamma_k=Ex_tx_{t+k} \qquad k=0,1,\cdots,m \qquad (10.1.14)$$

那么如何来估计 $g(\lambda)$ 呢?

我们知道，谱密度和自协方函数互相唯一确定（见 (1.3.7), (1.3.8) 式）. 只要 m 以后的 $\gamma_{m+1}, \gamma_{m+2}, \cdots$ 能使 $\gamma_k, k=0, 1, \cdots$ 成非负定列且 $\sum_{k=0}^{\infty}|\gamma_k|<\infty$，那么

$$g(\lambda)=\sum_{k=-\infty}^{\infty}\gamma_k e^{-i2\pi\lambda k}$$

必为一平稳列的谱密度，而且其前 $m+1$ 个相关函数值 $\gamma_0, \cdots, \gamma_m$ 就是预先已知的 $m+1$ 个数. 因此满足条件

$$\int_{-\frac{1}{2}}^{\frac{1}{2}}g(\lambda)e^{i2\pi\lambda k}\,d\lambda=\gamma_k, \quad k=0, \pm 1, \cdots, \pm m \qquad (10.1.15)$$

的 $g(\lambda)$ 是很多的. 我们应当这样来选择 $g_0(\lambda)$（真实谱）的估计，它所相应的序列是在满足条件 (10.1.15) 式条件下其随机性最强的. 因为除了 $\gamma_0, \cdots, \gamma_m$ 的值以外，其他毫无所知，所以除此约束以外的其他任何对随机性的约束都会带上不应有的主观因素. 正如前面指出的，以这样的 $g(\lambda)$ 为谱密度的序列是在满足

(10.1.14) 式的所有平稳序列中包含新息量最多的序列，它的谱熵也达极大.

在约束 (10.1.15) 式下，使得 (10.1.8) 式的谱熵 H 达极大值，以此作为准则所得到的估计称之为极大熵谱估计. 下面，按这一原则推导出估计量的具体表达式. 记

$$J(g)=\int_{-\frac{1}{2}}^{\frac{1}{2}}\log g(\lambda)d\lambda-\sum_{k=-m}^{m}\zeta_k\int_{-\frac{1}{2}}^{\frac{1}{2}}g(\lambda)e^{i2\pi\lambda k}d\lambda$$

$$=\int_{-\frac{1}{2}}^{\frac{1}{2}}\left[\log g(\lambda)-\left(\sum_{k=-m}^{m}\zeta_k e^{i2\pi\lambda k}\right)g(\lambda)\right]d\lambda$$

(10.1.16)

其中 ζ_k 是 Lagrange 乘子，它们都为实数. 根据 (10.1.15) 式 $\gamma_k=\gamma_{-k}$，所以我们可设

$$\zeta_{-k}=\zeta_k\quad(k=1,2,\cdots,m)\qquad(10.1.17)$$

使泛涵 $J(g)$ 达极值的条件是变分 $\delta J(g)=0$，和推导 (10.1.11) 式一样，(10.1.16) 式的变分为

$$\delta J(g)=\int_{-\frac{1}{2}}^{\frac{1}{2}}\left[\frac{1}{g(\lambda)}-\sum_{k=-m}^{m}\zeta_k e^{i2\pi\lambda k}\right]\delta g(\lambda)d\lambda$$

$\delta g(\lambda)$ 表在空间 $\mathscr{C}^0$ 中 $g(\lambda)$ 的任一个“增量”，于是 $\delta J(g)=0$ 等价于上式方括号中的内容等于零，即

$$g(\lambda)=\frac{1}{\sum_{k=-m}^{m}\zeta_k e^{i2\pi\lambda k}}\qquad(10.1.18)$$

记 $\zeta(z)=\sum_{k=-m}^{m}\zeta_k z^k$，用 z^* 表 z 之共轭复数，则

$$\zeta(z^*)=\sum_{k=-m}^{m}\zeta_k(z^*)^k=\left(\sum_{k=-m}^{m}\zeta_k z^k\right)^*=(\zeta(z))^*$$

又由 (10.1.17) 式，$\zeta(z^{-1})=\zeta(z)$，可见若某 z_k 是 $\zeta(z)$ 之零点，

则 z_k^*，z_k^{-1}，$(z_k^*)^{-1}$ 也都是 $\zeta(z)$ 之零点．另外，我们总设 $g(\lambda)$ 是可积的，即

$$\infty > \gamma_0 = \int_{-\frac{1}{2}}^{\frac{1}{2}} g(\lambda)d\lambda = \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{d\lambda}{\sum_{k=-m}^{m} \zeta_k e^{i2\pi\lambda k}} = \frac{1}{2\pi i}\oint_{|z|=1} \frac{dz}{z\zeta(z)}$$

所以 $\zeta(z)$ 不可能在单位圆上有零点．

记 $\zeta(z) = z^{-m}\bar{\zeta}(z)$，则 $\bar{\zeta}(z)$ 是 z 的 $2m$ 次多项式，它的零点和 $\zeta(z)$ 的零点完全一样，所以我们可以把 $\zeta(z)$ 写成

$$\begin{aligned}
\zeta(z) &= cz^{-m}\prod_{k=1}^{m}(z - z_k^*)(z - z_k^{-1}) \\
&= c\prod_{k=1}^{m}(z - z_k^*)(1 - z^{-1}z_k^{-1}) \\
&= c\left[\prod_{k=1}^{m}(-z_k)^{-1}\right]\left[\prod_{k=1}^{m}(z - z_k^*)(z^{-1} - z_k)\right] \\
&= c'\prod_{k=1}^{m}(z - z_k^*)(z^{-1} - z_k)
\end{aligned}$$

其中 c, c' 都表示常数．于是

$$\begin{aligned}
g(\lambda) &= \frac{1}{\zeta(e^{i2\pi\lambda})} = \frac{1}{c'\prod_{k=1}^{m}(e^{i2\pi\lambda} - z_k^*)(e^{-i2\pi\lambda} - z_k)} \\
&= \frac{1}{c'\prod_{k=1}^{m}(1 - z_k^* e^{-i2\pi\lambda})(1 - z_k e^{i2\pi\lambda})} \\
&= \frac{1}{c'\left|\prod_{k=1}^{m}(1 - z_k e^{i2\pi\lambda})\right|^2}
\end{aligned} \tag{10.1.19}$$

记

$$\prod_{k=1}^{m}(1 - z_k e^{i2\pi\lambda}) = 1 - \varphi_1 e^{i2\pi\lambda} - \cdots - \varphi_m e^{i2\pi\lambda m} = \varphi(e^{i2\pi\lambda})$$

前面已指出 z_k，z_k^*，z_k^{-1}，$(z_k^*)^{-1}$ 都是 $\zeta(z)$ 的零点，所以我们可以

把 $z_1, \cdots, z_m$ 都取为单位圆内的 m 个零点，显见它们或者是实数，或者成共轭地存在于这组 $z_1, \cdots, z_m$ 中，因而可使 $\varphi_1, \cdots, \varphi_m$ 都是实数，且 $\varphi(z)=\prod_{k=1}^{m}(1-z_k z)$ 的零点 z_k^{-1} 全部在单位圆外.

另外，由 $g(\lambda)>0$，知 $c'>0$，故可记 $\frac{1}{c'}=\sigma^2$. 因此，最后我们将极大熵谱估计 $g(\lambda)$ 写成

$$g(\lambda)=\sigma^2/\varphi(e^{i2\pi\lambda})\varphi^*(e^{i2\pi\lambda})=\sigma^2/|\varphi(e^{i2\pi\lambda})|^2 \tag{10.1.20}$$

这实际上是 m 阶的 AR 序列的谱密度.

这就证明了，存在一个 m 阶的 AR 序列，其前 $m+1$ 个自协方函数值等于已知值（见(10.1.15)式），而使谱熵(参见(10.1.8)式)达极大值.

我们还须给出 $\varphi(z)$ 的系数的表达式.

在(10.1.20)式中，两端乘以 $\varphi^*(e^{i2\pi\lambda})e^{i2\pi\lambda k}=\varphi(e^{-i2\pi\lambda})e^{i2\pi\lambda k}$ 并积分 $(k=0, 1, \cdots, m)$，则由(10.1.15)式，左端等于

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} g(\lambda)\varphi^*(e^{i2\pi\lambda})e^{i2\pi\lambda k}d\lambda=\gamma_k-\varphi_1\gamma_{k-1}-\cdots-\varphi_m\gamma_{k-m} \tag{10.1.21}$$

右端为

$$\sigma^2\int_{-\frac{1}{2}}^{\frac{1}{2}}\frac{e^{i2\pi\lambda k}}{\varphi(e^{i2\pi\lambda})}d\lambda=\frac{\sigma^2}{2\pi i}\oint_{|z|=1}\frac{z^{k-1}}{\varphi(z)}dz=\sigma^2\delta_{k,0}$$

$$k=0, 1, \cdots, m \tag{10.1.22}$$

因为 $\varphi(z)$ 在单位圆内没有零点，$z^{k-1}/\varphi(z)$ 只有在 $k=0$ 时留数等于 $\frac{1}{\varphi(0)}=1$，在其他情况留数为零. 从(10.1.21)，(10.1.22)式得

$$\gamma_k-\varphi_1\gamma_{k-1}-\cdots-\varphi_m\gamma_{k-m}=\sigma^2\delta_{k,0},$$

$$k=0, 1, \cdots, m \tag{10.1.23}$$

注意 $\gamma_{-k}=\gamma_k$，上式写成矩阵形式为

$$
\begin{pmatrix} \gamma_0 \cdots \gamma_m \\ \cdots\cdots \\ \gamma_m \cdots \gamma_0 \end{pmatrix} \begin{pmatrix} 1 \\ -\varphi_1 \\ \vdots \\ -\varphi_m \end{pmatrix} = \begin{pmatrix} \sigma^2 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \qquad (10.1.24)
$$

取其后 m 个方程，即

$$
\begin{pmatrix} \gamma_0 \cdots \gamma_{m-1} \\ \cdots\cdots\cdots \\ \gamma_{m-1} \cdots \gamma_0 \end{pmatrix} \begin{pmatrix} \varphi_1 \\ \vdots \\ \varphi_m \end{pmatrix} = \begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_m \end{pmatrix} \qquad (10.1.25)
$$

这正是 Yule-Walker 方程，由它解出 $\varphi_1, \cdots, \varphi_m$，代入 (10.1.24) 式中第一个方程可得 σ^2.

总之，知道了前 $m+1$ 个自协方函数值以后，极大熵谱估计便是按 (10.1.25) 式 Yule-Walker 方程求得的 m 阶 AR 模型的谱.

§2. 极大熵谱估计的计算方法

如果确切地知道前 $m+1$ 个自协方函数值，用前一节中 (10.1.24),(10.1.20) 式可求得具有极大谱熵的 AR 序列的参数.而 (10.1.25)式中 $\varphi_1, \cdots, \varphi_m$ 的解法可利用第二章 §2 中求偏相关函数的递推公式求得. 但是在实际问题中，只知道 n 个样本 $x_1, \cdots, x_n$，m 的大小并不知道，而 $\gamma_0, \cdots, \gamma_{\pm m}$ 又只能用估计代替.

通常给出的 γ_k 的估计量有下列两种

$$
\hat{\gamma}'_k = \frac{1}{n}\sum_{t=1}^{n-k} x_t x_{t+k} \qquad (10.2.1)
$$

$$
\hat{\gamma}''_k = \frac{1}{n-k}\sum_{t=1}^{n-k} x_t x_{t+k} \qquad (10.2.2)
$$

从偏差上来说，对较大的 k(尤其是 k 接近 n 时)，仍有 $E\hat{\gamma}''_k = \gamma_k$，而 $\hat{\gamma}'_k$ 的误差偏差却较大，但 $\gamma''_0, \gamma''_1, \cdots, \gamma''_m$ 不能保证样本自协方差阵

$$
\begin{pmatrix} \hat{\gamma}''_0 \cdots \hat{\gamma}''_m \\ \cdots\cdots \\ \hat{\gamma}''_m \cdots \hat{\gamma}''_0 \end{pmatrix} \qquad (m \leqslant n-1)
$$

的非负定性. 所以用(10.2.1), (10.2.2)式作为(10.1.24)式中 γ_k 的估计各有不足之处. 因此在应用上节得到的准则时，引入了一种新的方法.

在引进新方法之前，先介绍一下所谓 Wiggins-Robinson 辅助系数迭代法.

设 $\gamma_0, \gamma_1, \cdots$ 是正定列. 对一切 n，令 $\phi_{n0}=1$；而 $n+1$ 个未知量 $\phi_{n1}, \cdots, \phi_{nn}, \alpha_n$ 满足下列方程组

$$\begin{pmatrix} \gamma_0 \cdots \gamma_n \\ \cdots\cdots \\ \gamma_n \cdots \gamma_0 \end{pmatrix} \begin{pmatrix} \phi_{n0} \\ \vdots \\ \phi_{nn} \end{pmatrix} = \begin{pmatrix} \alpha_n \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{10.2.3}$$

显然

$$\alpha_n = (\phi_{n0}, \cdots, \phi_{nn}) \begin{pmatrix} \alpha_n \\ 0 \\ \vdots \\ 0 \end{pmatrix} = (\phi_{n0}, \cdots, \phi_{nn}) \begin{pmatrix} \gamma_0 \cdots \gamma_n \\ \cdots\cdots \\ \gamma_n \cdots \gamma_0 \end{pmatrix} \begin{pmatrix} \phi_{n0} \\ \vdots \\ \phi_{nn} \end{pmatrix} > 0$$

$$\begin{pmatrix} \gamma_0 \cdots \gamma_n \\ \cdots\cdots \\ \gamma_n \cdots \gamma_0 \end{pmatrix} \begin{pmatrix} \phi_{nn} \\ \vdots \\ \phi_{n0} \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \alpha_n \end{pmatrix} \tag{10.2.4}$$

现考虑 $n+2$ 个未知数 $\phi_{n+1,1}, \cdots \phi_{n+1,n+1}, \alpha_{n+1}$ 关于方程组

$$\begin{pmatrix} \gamma_0 \cdots \gamma_{n+1} \\ \cdots\cdots \\ \gamma_{n+1} \cdots \gamma_0 \end{pmatrix} \begin{pmatrix} \phi_{n+1,0} \\ \vdots \\ \phi_{n+1,n+1} \end{pmatrix} = \begin{pmatrix} \alpha_{n+1} \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{10.2.5}$$

的解. 我们可由(10.2.3), (10.2.4)式分别得

$$\begin{pmatrix} \gamma_0 \cdots \gamma_{n+1} \\ \cdots\cdots\cdots \\ \gamma_{n+1} \cdots \gamma_0 \end{pmatrix} \begin{pmatrix} \phi_{n0} \\ \vdots \\ \phi_{nn} \\ 0 \end{pmatrix} = \begin{pmatrix} \alpha_n \\ 0 \\ \vdots \\ 0 \\ \alpha_n' \end{pmatrix} \tag{10.2.6}$$

$$
\begin{pmatrix} \gamma_0 \cdots \gamma_{n+1} \\ \cdots\cdots\cdots \\ \gamma_{n+1} \cdots \gamma_0 \end{pmatrix} \begin{pmatrix} 0 \\ \phi_{nn} \\ \vdots \\ \phi_{n0} \end{pmatrix} = \begin{pmatrix} \alpha'_n \\ 0 \\ \vdots \\ 0 \\ \alpha_n \end{pmatrix} \tag{10.2.7}
$$

其中，α'_n 很容易计算出来．取 k_n，使它满足（由 $\alpha_n > 0$，下面方程有唯一解）

$$
\alpha'_n + k_n \alpha_n = 0 \tag{10.2.8}
$$

暂时令

$$
\begin{aligned}
&(\phi_{n0}, \cdots, \phi_{nn}, 0)^\tau + k_n(0, \phi_{nn}, \cdots, \phi_{n0})^\tau \\
&= (\phi_{n+1,0}, \cdots, \phi_{n+1,n+1})^\tau
\end{aligned} \tag{10.2.9}
$$

则由 (10.2.6), (10.2.7) 式

$$
\begin{pmatrix} \gamma_0 \cdots \gamma_{n+1} \\ \cdots\cdots \\ \gamma_{n+1} \cdots \gamma_0 \end{pmatrix} \begin{pmatrix} \phi_{n+1,0} \\ \vdots \\ \phi_{n+1,n+1} \end{pmatrix} = \begin{pmatrix} \alpha_n \\ 0 \\ \vdots \\ 0 \\ \alpha'_n \end{pmatrix} + k_n \begin{pmatrix} \alpha'_n \\ 0 \\ \vdots \\ 0 \\ \alpha_n \end{pmatrix} = \begin{pmatrix} \alpha_n + k_n \alpha'_n \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{10.2.10}
$$

显见 $\phi_{n+1,0} = 1$，即由 (10.2.9) 式给出的 $(\phi_{n+1,1}, \cdots, \phi_{n+1,n+1})$ 和 $\alpha_{n+1} = \alpha_n + k_n \alpha'_n$ 是 (10.2.5) 式的解．由此，可逐步迭代求任意阶的形如 (10.2.3) 式的方程组之解．其中 $(\phi_{n0}, \cdots, \phi_{nn})$ 就称为 *W-R* 辅助系数．

我们知道，m 阶的 AR 模型相当于一个时间从 $-\infty$ 开始的，带有直至 m 步延迟反馈的线性滤波器，序列 x_t 本身就是以白噪声 a_t 为输入的此滤波器的输出：

$$
x_t = \varphi_1 x_{t-1} + \cdots + \varphi_m x_{t-m} + a_t \tag{10.2.11}
$$

输入 a_t 是用 t 时刻前的输出来预报 t 时刻输出 x_t（线性最小方差预报）的预报误差．

现若把序列 x_t 从原输出端反输入到该滤波器中，并在滤波器内部把全部流向都倒过来，那么，就在原输入端得到输出，若时间从 $-\infty$ 开始，这个输出就是 a_t：

$$a_t = x_t - \varphi_1 x_{t-1} - \cdots - \varphi_m x_{t-m} \tag{10.2.12}$$

因此，我们将原来滤波器倒过来用时，可称作"预报误差滤波器".

下面，我们给出不同于(10.2.1)，(10.2.2)式的估计 $\hat{\gamma}_k$ 的新方法，同时给出求极大熵谱估计中诸参数的迭代程序．在一些文献中称为 Burg 方法．

设 $Ex_t = 0$，把m阶预报误差滤波器（见(10.2.12)式）的输出的方差记为 σ_m^2，而其中相应的系数记为 $\varphi_{m1}, \cdots, \varphi_{mm}$.

令 $m = 0$，则

$$\sigma_0^2 = \frac{1}{n}\sum_{t=1}^{n} x_t^2 = \hat{\gamma}_0 > 0 \tag{10.2.13}$$

是零阶预报误差滤波器输出方差的估计.

令 $m = 1$，按极大熵谱估计准则推导出的公式（10.1.24)式，并用 $\hat{\gamma}_k$ 代替其中 γ_k，有

$$\begin{pmatrix}\sigma_1^2\\0\end{pmatrix} = \begin{pmatrix}\hat{\gamma}_0 & \hat{\gamma}_1\\ \hat{\gamma}_1 & \hat{\gamma}_0\end{pmatrix}\begin{pmatrix}1\\-\varphi_{11}\end{pmatrix} = \begin{pmatrix}\hat{\gamma}_0 & \hat{\gamma}_1\\ \hat{\gamma}_1 & \hat{\gamma}_0\end{pmatrix}\left[\begin{pmatrix}1\\0\end{pmatrix} - \varphi_{11}\begin{pmatrix}0\\1\end{pmatrix}\right]$$

$$= \begin{pmatrix}\sigma_0^2\\ \hat{\gamma}_1\end{pmatrix} - \varphi_{11}\begin{pmatrix}\hat{\gamma}_1\\ \sigma_0^2\end{pmatrix}$$

得

$$\hat{\gamma}_1 = \varphi_{11}\sigma_0^2 \tag{10.2.14}$$

$$\sigma_1^2 = \sigma_0^2 - \varphi_{11}\hat{\gamma}_1 = (1 - \varphi_{11}^2)\sigma_0^2 \tag{10.2.15}$$

其中，φ_{11} 应使一阶预报误差滤波器的输出方差的估计量

$$\frac{1}{2(n-1)}\left[\sum_{t=2}^{n}(x_t - \varphi_{11}x_{t-1})^2 + \sum_{t=2}^{n}(x_{t-1} - \varphi_{11}x_t)^2\right]$$

达极小．其中后一和号相当于把数据 $x_1, \cdots, x_n$ 再倒过来使用一次以提高数据利用率，因为 $(x_n, \cdots, x_1)$ 和 $(x_1, \cdots, x_n)$ 有完全相同的统计特性．于是 φ_{11} 满足方程

$$\frac{\partial}{\partial\varphi_{11}}\left[\sum_{t=2}^{n}\left[(x_t - \varphi_{11}x_{t-1})^2 + (x_{t-1} - \varphi_{11}x_t)^2\right]\right] = 0$$

即

$$\varphi_{11} = 2\sum_{t=2}^{n} x_t x_{t-1} \Big/ \left(\sum_{t=2}^{n} x_t^2 + \sum_{t=2}^{n} x_{t-1}^2\right) \tag{10.2.16}$$

令 $m = 2$，根据求 $W-R$ 辅助系数的思想，我们有

$$\begin{pmatrix} \sigma_2^2 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} \hat{\gamma}_0 & \hat{\gamma}_1 & \hat{\gamma}_2 \\ \hat{\gamma}_1 & \hat{\gamma}_0 & \hat{\gamma}_1 \\ \hat{\gamma}_2 & \hat{\gamma}_1 & \hat{\gamma}_0 \end{pmatrix} \begin{pmatrix} 1 \\ -\varphi_{21} \\ -\varphi_{22} \end{pmatrix}$$

$$= \begin{pmatrix} \hat{\gamma}_0 & \hat{\gamma}_1 & \hat{\gamma}_2 \\ \hat{\gamma}_1 & \hat{\gamma}_0 & \hat{\gamma}_1 \\ \hat{\gamma}_2 & \hat{\gamma}_1 & \hat{\gamma}_0 \end{pmatrix} \left[\begin{pmatrix} 1 \\ -\varphi_{11} \\ 0 \end{pmatrix} - \varphi_{22} \begin{pmatrix} 0 \\ -\varphi_{11} \\ 1 \end{pmatrix} \right]$$

$$= \begin{pmatrix} \sigma_1^2 \\ 0 \\ \hat{\gamma}_2 - \varphi_{11}\hat{\gamma}_1 \end{pmatrix} - \varphi_{22} \begin{pmatrix} \hat{\gamma}_2 - \varphi_{11}\hat{\gamma}_1 \\ 0 \\ \sigma_1^2 \end{pmatrix}$$

得

$$\hat{\gamma}_2 = \varphi_{11}\hat{\gamma}_1 + \varphi_{22}\sigma_1^2 \tag{10.2.17}$$

$$\sigma_2^2 = \sigma_1^2 - \varphi_{22}(\hat{\gamma}_2 - \varphi_{11}\hat{\gamma}_1) = (1 - \varphi_{22}^2)\sigma_1^2 \tag{10.2.18}$$

$$\varphi_{21} = \varphi_{11} - \varphi_{22}\varphi_{11} \tag{10.2.19}$$

其中，φ_{22} 应当使二阶预报误差滤波器的输出方差的估计量

$$\frac{1}{2(n-2)}\sum_{t=3}^{n}\{[x_t - \varphi_{21}x_{t-1} - \varphi_{22}x_{t-2}]^2 + [x_{t-2} - \varphi_{21}x_{t-1} - \varphi_{22}x_t]^2\}$$

$$= \frac{1}{2(n-2)}\sum_{t=3}^{n}\{[(x_t - \varphi_{11}x_{t-1}) - \varphi_{22}(x_{t-2} - \varphi_{11}x_{t-1})]^2 + [(x_{t-2} - \varphi_{11}x_{t-1}) - \varphi_{22}(x_t - \varphi_{11}x_{t-1})]^2\}$$

达极小，即由此式对 φ_{22} 求偏导数等于零解出 φ_{22} 得

$$\varphi_{22} = \frac{2\sum_{t=3}^{n}(x_t - \varphi_{11}x_{t-1})(x_{t-2} - \varphi_{11}x_{t-1})}{\sum_{t=3}^{n}[(x_t - \varphi_{11}x_{t-1})^2 + (x_{t-2} - \varphi_{11}x_{t-1})^2]} \tag{10.2.20}$$

若对于 $j \leqslant m$ 已得到 $\hat{\gamma}_m$, σ_m^2, φ_{mj}，现在求出用它们递推 $\hat{\gamma}_{m+1}$, σ_{m+1}^2, $\varphi_{m+1,j}$ $(j \leqslant m+1)$ 的公式．同样根据求 W-R 辅助系数的思想，考虑 $m+2$ 个联立方程组

$$
\begin{pmatrix} \sigma_{m+1}^2 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} \hat{\gamma}_0 & \cdots & \hat{\gamma}_{m+1} \\ & \ddots & \\ \hat{\gamma}_{m+1} & \cdots & \hat{\gamma}_0 \end{pmatrix} \begin{pmatrix} 1 \\ -\varphi_{m+1,1} \\ \vdots \\ -\varphi_{m+1,m+1} \end{pmatrix}
$$

$$
= \begin{pmatrix} \hat{\gamma}_0 & \cdots & \hat{\gamma}_{m+1} \\ & \ddots & \\ \hat{\gamma}_{m+1} & \cdots & \hat{\gamma}_0 \end{pmatrix} \left[\begin{pmatrix} 1 \\ -\varphi_{m1} \\ \vdots \\ -\varphi_{mm} \\ 0 \end{pmatrix} - \varphi_{m+1,m+1} \begin{pmatrix} 0 \\ -\varphi_{mm} \\ \vdots \\ -\varphi_{m1} \\ 1 \end{pmatrix} \right]
$$

$$
= \begin{pmatrix} \sigma_m^2 \\ 0 \\ \vdots \\ 0 \\ \hat{\gamma}_{m+1} - \sum_{j=1}^{m} \varphi_{mj} \hat{\gamma}_{m+1-j} \end{pmatrix} - \varphi_{m+1,m+1} \begin{pmatrix} \hat{\gamma}_{m+1} - \sum_{j=1}^{m} \varphi_{mj} \hat{\gamma}_{m+1-j} \\ 0 \\ \vdots \\ 0 \\ \sigma_m^2 \end{pmatrix} \tag{10.2.21}
$$

得

$$
\hat{\gamma}_{m+1} = \sum_{j=1}^{m} \varphi_{mj} \hat{\gamma}_{m+1-j} + \varphi_{m+1,m+1} \sigma_m^2 \tag{10.2.22}
$$

$$
\begin{aligned}
\sigma_{m+1}^2 &= \sigma_m^2 - \varphi_{m+1,m+1} \left(\hat{\gamma}_{m+1} - \sum_{j=1}^{m} \varphi_{mj} \hat{\gamma}_{m+1-j} \right) \\
&= (1 - \varphi_{m+1,m+1}^2) \sigma_m^2
\end{aligned} \tag{10.2.23}
$$

$$
\varphi_{m+1,j} = \varphi_{mj} - \varphi_{m+1,m+1} \varphi_{m,m+1-j} \tag{10.2.24}
$$

其中 $\varphi_{m+1,m+1}$ 应使 $m+1$ 阶预报误差滤波器输出方差的估计量达

$$\frac{1}{2(n-m-1)}\sum_{t=m+2}^{n}\left\{\left[x_t-\sum_{j=1}^{m+1}\varphi_{m+1,j}x_{t-j}\right]^2\right.$$

$$\left.+\left[x_{t-m-1}-\sum_{j=1}^{m+1}\varphi_{m+1,j}x_{t-m-1+j}\right]^2\right\}$$

$$=\frac{1}{2(n-m-1)}\sum_{t=m+2}^{n}\left\{\left[\left(x_t-\sum_{j=1}^{m}\varphi_{mj}x_{t-j}\right)\right.\right.$$

$$\left.-\varphi_{m+1,m+1}\left(x_{t-m-1}-\sum_{j=1}^{m}\varphi_{m,m+1-j}x_{t-j}\right)\right]^2$$

$$+\left[\left(x_{t-m-1}-\sum_{j=1}^{m}\varphi_{m,j}x_{t-m-1-j}\right)\right.$$

$$\left.\left.-\varphi_{m+1,m+1}\left(x_t-\sum_{j=1}^{m}\varphi_{mj}x_{t-m-1+j}\right)\right]^2\right\}$$

极小，通过对 $\varphi_{m+1,m+1}$ 偏导数等于0，得

$$\varphi_{m+1,m+1}=\left[2\sum_{t=m+2}^{n}\left(x_t-\sum_{j=1}^{m}\varphi_{mj}x_{t-j}\right)\right.$$

$$\left.\cdot\left(x_{t-m-1}-\sum_{j=1}^{m}\varphi_{mj}x_{t-m-1+j}\right)\right]\Big/$$

$$\left\{\sum_{t=m+2}^{n}\left[\left(x_t-\sum_{j=1}^{m}\varphi_{mj}x_{t-j}\right)^2\right.\right.$$

$$\left.\left.+\left(x_{t-m-1}-\sum_{j=1}^{m}\varphi_{mj}x_{t-m-1+j}\right)^2\right]\right\} \qquad (10.2.25)$$

(10.2.22) 至 (10.2.25) 式是一般的递推公式. 根据全部数据以及前一步的系数 φ_{mj}，用 (10.2.25) 式求得 $\varphi_{m+1,m+1}$；由 (10.2.24) 式得 $\varphi_{m+1,j}$ $(j<m+1)$；(10.2.23) 式给出了该步预报误差滤波器的输出方差的估计量. (10.2.22) 式实际上给出了不同于 (10.2.1)，(10.2.2) 式那样通常的 γ_k 的估计公式.

现在论证上述的 $\hat{\gamma}_0$, $\hat{\gamma}_1$, $\cdots$ 是正定列. 首先由(10.2.25)式知

$$
\begin{aligned}
|\varphi_{m+1,m+1}| \leqslant & \left[2 \sum_{t=m+2}^{n} \left| \left(x_t - \sum_{j=1}^{m} \varphi_{mj} x_{t-j} \right) \right.\right. \\
& \left.\left. \cdot \left(x_{t-m-1} - \sum_{j=1}^{m} \varphi_{mj} x_{t-m-1+j} \right) \right| \right] \Big/ \\
& \left\{ \sum_{t=m+2}^{n} \left[\left(x_t - \sum_{j=1}^{m} \varphi_{mj} x_{t-j} \right)^2 \right.\right. \\
& \left.\left. + \left(x_{t-m-1} - \sum_{j=1}^{m} \varphi_{mj} x_{t-m-1+j} \right)^2 \right] \right\} \\
& \leqslant \left\{ \sum_{t=m+2}^{n} \left[\left(x_t - \sum_{j=1}^{m} \varphi_{mj} x_{t-j} \right)^2 + \left(x_{t-m+1} \right.\right.\right. \\
& \left.\left.\left. - \sum_{j=1}^{m} \varphi_{mj} x_{t-m-1+j} \right)^2 \right] \right\} \Big/ \left\{ \sum_{t=m+2}^{n} \left[\left(x_t - \right.\right.\right. \\
& \left.\left.\left. - \sum_{j=1}^{m} \varphi_{mj} x_{t-j} \right)^2 + \left(x_{t-m+1} - \sum_{j=1}^{m} \varphi_{mj} x_{t-m-1+j} \right)^2 \right] \right\} \\
& \leqslant 1
\end{aligned}
$$

以此代入(10.2.23)式又知 $0 \leqslant \sigma_{m+1}^2 \leqslant \sigma_m^2$; 再由(10.2.21)式的第一等式知

$$
|\hat{\Gamma}_{m+1}| / |\hat{\Gamma}_m| = \sigma_{m+1}^2 \geqslant 0 \qquad (10.2.26)
$$

其中 $\hat{\Gamma}_{m+1}$ 就是(10.2.21)式中那个由 $\hat{\gamma}_k$ 构成的 $m+2$ 阶方阵, $\hat{\Gamma}_m$ 是其$m+1$ 阶主子阵. 再注意 $|\hat{\Gamma}_0| = \sigma_0^2 > 0$, 于是,根据归纳法, 若 $\hat{\Gamma}_k > 0$ $(k = 0, 1, \cdots, m)$, 则由(10.2.26)式知必有 $|\hat{\Gamma}_{m+1}| > 0$ (且(10.2.26)式中 >0 成立). 由此可见用这种方法迭代计算出的 $\hat{\gamma}_0, \hat{\gamma}_1, \cdots$ 构成正定列. 至于它们的渐近均值、渐近均方误差及渐近分布等问题,尚未进行研究,但由于用它作谱估计时效果较好,所以在实际中已被采用.

最后谈一下选阶的问题. 如果序列本身是一无穷阶的自回归模型,从理论上讲,当数据个数为 n 时,选用 $n-1$ 阶最好. 但是由于m 增大时,用 $\hat{\gamma}_m$ 估计 γ_m 变得越来越不准, 所以取$m = n-1$

未必最好，尤其当 n 较大时更没有必要，而且计算量也太大．一般只要取 m 为 n 的若干分之一，如 $n/2$，$n/3$ 等等就可以了．

如果序列本来是一个 p 阶自回归，那么应当取 m 尽可能接近 p．为此，在作谱估计之前可先用偏相关函数检验一下序列是否可看成有限阶自回归．如果能，且阶数 $m \ll n$（如 $m < 0.1n$），那么偏相关系数 $\varphi_{1m}, \cdots, \varphi_{mm}$ 就可作为其参数估计，用它和用 Burg 方法得到的结果相差不会太大．

以上关于选择 m 值的方法（如取 $n/2$，$n/3$，$0.1n$ 等）都是经验方法，更严格的选定 m 值的数学方法，还在探索之中．

附　　录

§1. 自相关与偏相关函数的截尾性质

随机序列 w_t 称为 ARMA (p, q) 序列，如果它是随机模型

$$w_t - \varphi_1 w_{t-1} - \cdots - \varphi_p w_{t-p} = a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q} \tag{1.1}$$

的平稳解，其中 $\varphi(\omega) = 1 - \varphi_1\omega - \varphi_2\omega^2 - \cdots - \varphi_p\omega^p$，$\theta(\omega) = 1 - \theta_1\omega - \cdots - \theta_q\omega^q$ 的根都在单位圆外，a_t 是白噪声序列，$Ea_t = 0$，$Ea_t^2 = \sigma^2$，$Ea_t\omega_s = 0 (s < t)$，本节暂不假定它是正态序列. 我们将讨论，一个平稳随机序列在什么条件下成为 ARMA (p, q) 序列，又在什么条件下成为 AR $(p, 0)$ 序列或 MA $(0, q)$ 序列.

本节各定理证明的主要依据是宽平稳随机序列的谱表示定理（见[2]X§4），下面简称谱表示定理.

定理 1. 均值为零的平稳序列 w_t 满足随机差分方程(1.1)的充分必要条件是，w_t 具有有理谱密度

$$f(\lambda) = \frac{|\theta(e^{-i2\pi\lambda})|^2}{|\varphi(e^{-i2\pi\lambda})|^2}\sigma_a^2 \tag{1.2}$$

证明：先证必要性.

根据谱表示定理，可将(1.1)式中的平稳白噪声序列 a_t 表成

$$a_t = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{i2\pi t\lambda} dy_a(\lambda) \tag{1.3}$$

其中 $y_a(\lambda)$ 是相应于 a_t 的正交增量过程，

$$E\{|dy_a(\lambda)|^2\} = \sigma_a^2 d\lambda$$

又根据同一定理可知

$$w_t = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{i2\pi t\lambda} dy_w(\lambda) \tag{1.4}$$

其中 $y_w(\lambda)$ 是相应于平稳序列 w_t 的正交增量过程，

$$E\{|dy_w(\lambda)|^2\} = dF(\lambda)$$

其中 $F(\lambda)$ 是 w_t 的谱分布函数.由于已经假定了 w_t 满足(1.1)式，所以将(1.3)和(1.4)式代入(1.1)式可得

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} e^{i2\pi t\lambda}\varphi(e^{-i2\pi\lambda})dy_w(\lambda) = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{i2\pi t\lambda}\theta(e^{-i2\pi\lambda})dy_a(\lambda)$$

再用谱表示定理可知,上式两边的随机序列的自协方差函数 γ_k 分别是下式两边的数列，

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} e^{i2\pi k\lambda}|\varphi(e^{-i2\pi\lambda})|^2dF(\lambda) = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{i2\pi k\lambda}|\theta(e^{-i2\pi\lambda})|^2\sigma_a^2d\lambda$$

它对一切整数 k 都成立,因此

$$|\varphi(e^{-i2\pi\lambda})|^2dF(\lambda) = |\theta(e^{-i2\pi\lambda})|^2\sigma_a^2d\lambda$$

又由于 $\varphi(e^{-i2\pi\lambda}) \neq 0$（因为 $|e^{-i2\pi\lambda}| = 1$），所以

$$\frac{dF(\lambda)}{d\lambda} = f(\lambda) = \frac{|\theta(e^{-i2\pi\lambda})|^2}{|\varphi(e^{-i2\pi\lambda})|^2}\sigma_a^2$$

再证充分性. 若已知 w_t 具有(1.2)式的有理谱密度,那么

$$\gamma_k = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{i2\pi k\lambda}f(\lambda)d\lambda = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{i2\pi k\lambda}\frac{|\theta(e^{-i2\pi\lambda})|^2}{|\varphi(e^{-i2\pi\lambda})|^2}\sigma_a^2d\lambda$$

由谱表示定理可知

$$w_t = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{i2\pi t\lambda}\frac{\theta(e^{-i2\pi\lambda})}{\varphi(e^{-i2\pi\lambda})}dy(\lambda)$$

其中 $y(\lambda)$ 为正交增量过程，$E\{|dy(\lambda)|^2\} = \sigma_a^2d\lambda$. 由此便可得到

$$w_t - \varphi_1w_{t-1} - \cdots - \varphi_pw_{t-p} = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{i2\pi t\lambda}\theta(e^{-i2\pi\lambda})dy(\lambda)$$
$$= a_t - \theta_1a_{t-1} - \cdots - \theta_qa_{t-q}$$

其中

$$a_t = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{i2\pi t\lambda}dy(\lambda)$$

$$Ea_t\bar{a}_s = \sigma_a^2\int_{-\frac{1}{2}}^{\frac{1}{2}} e^{i2\pi(t-s)\lambda}d\lambda = \begin{cases}\sigma_a^2, & t = s\\ 0, & t \neq s\end{cases}$$

故 a_t 为白噪声序列. 定理证明完毕.

定理 2. 设 w_t 为一零均值的平稳随机序列，它的自协方差函数 γ_k 是 q 步截尾的(即 $\gamma_q \neq 0$; $\gamma_k = 0, k > q$). 若进一步设函

数 $f(\lambda)=\sum_{k=-q}^{q}\gamma_k e^{-i2\pi k\lambda}\neq 0$，$-\frac{1}{2}\leqslant\lambda\leqslant\frac{1}{2}$，则 w_t 一定是 MA$(0, q)$ 序列.

证明. 首先有

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} e^{i2\pi j\lambda}f(\lambda)d\lambda=\sum_{k=-\infty}^{\infty}\gamma_k\int_{-\frac{1}{2}}^{\frac{1}{2}} e^{i2\pi(j-k)}d\lambda=\gamma_j$$

$$j=0,\ \pm 1,\ \pm 2,\ \cdots$$

又当 $N>q$ 时，

$$0\leqslant E\left|\sum_{j=1}^{N} w_j e^{-i2\pi j\lambda}\right|^2=\sum_{j,\,l=1}^{N}\gamma_{j-l}e^{-i2\pi(j-l)\lambda}$$

$$=\sum_{k=-N+1}^{N-1}(N-|k|)\gamma_k e^{-i2\pi k\lambda}$$

$$=N\sum_{k=-q}^{q}\left(1-\frac{|k|}{N}\right)\gamma_k e^{-i2\pi k\lambda}$$

所以

$$f(\lambda)=\lim_{N\to\infty}\sum_{k=-q}^{q}\left(1-\frac{|k|}{N}\right)\gamma_k e^{-i2\pi k\lambda}\geqslant 0$$

由此知 w_t 的谱密度存在，且为 $f(\lambda)$. 将 $f(\lambda)$ 表为

$$f(\lambda)=e^{i2\pi q\lambda}\sum_{k=-q}^{q}\gamma_k e^{-i2\pi(k+q)\lambda}=e^{i2\pi q\lambda}\sum_{j=0}^{2q}\gamma_{j-q}(e^{-i2\pi\lambda})^j$$

$$=\gamma_q e^{i2\pi q\lambda}\prod_{j=1}^{2q}(e^{-i2\pi\lambda}-\xi_j)$$

其中 ξ_j 是 $2q$ 次多项式

$$\sum_{j=0}^{2q}\gamma_{j-q}w^j=\gamma_q\prod_{j=1}^{2q}(w-\xi_j) \tag{1.5}$$

的根. 由于 $\gamma_{-q}=\gamma_q\neq 0$，所以 $\prod_{j=1}^{2q}\xi_j=1$，从而 $\xi_j\neq 0(1\leqslant$

$j\leqslant 2q$)；又由于 $f(\lambda)$ 是非负实函数，故 $\bar{f}(\lambda)=f(\lambda)$，即

$$\gamma_q e^{i2\pi q\lambda}\prod_{j=1}^{2q}(e^{-i2\pi\lambda}-\xi_j)=\gamma_q e^{-i2\pi q\lambda}\prod_{j=1}^{2q}(e^{i2\pi\lambda}-\bar{\xi}_j)$$

$$=\gamma_q e^{i2\pi q\lambda}\prod_{j=1}^{2q}\left(\frac{1}{\bar{\xi}_j}-e^{-i2\pi\lambda}\right)$$

其中 $\bar{f}(\lambda)$和 $\bar{\xi}_j$ 分别表示$f(\lambda)$和 ξ_j 的共轭复数．由上式可知，$1/\bar{\xi}_j$ 也是 (1.5) 式多项式的根，因此不妨假定它的 $2q$ 个根中后 q 个根恰好是前 q 个根的共轭倒数．由此可表

$$f(\lambda)=\gamma_q e^{i2\pi q\lambda}\prod_{j=1}^{q}(e^{-i2\pi\lambda}-\xi_j)\prod_{j=1}^{q}\left(e^{-i2\pi\lambda}-\frac{1}{\bar{\xi}_j}\right)$$

$$=\gamma_q e^{i2\pi q\lambda}\prod_{j=1}^{q}(e^{-i2\pi\lambda}-\xi_j)\left(e^{-i2\pi\lambda}-\frac{1}{\bar{\xi}_j}\right)$$

再利用 $f(\lambda)$ 的非负性可得

$$f(\lambda)=|f(\lambda)|=|\gamma_q|\prod_{j=1}^{q}\left|(e^{-i2\pi\lambda}-\xi_j)\left(e^{-i2\pi\lambda}-\frac{1}{\bar{\xi}_j}\right)\right|$$

$$=|\gamma_q|\prod_{j=1}^{q}|\xi_j|\prod_{j=1}^{q}|(1-\xi_j^{-1}e^{-i2\pi\lambda})||(1-\bar{\xi}_j^{-1}e^{i2\pi\lambda})|$$

$$=\sigma^2\left|\prod_{j=1}^{q}(1-\xi_j^{-1}e^{-i2\pi\lambda})\right|^2$$

$$=\sigma^2\left|\sum_{j=0}^{q}\zeta_j e^{-i2\pi\lambda j}\right|^2$$

其中 $\sigma^2=|\gamma_q|\prod_{j=1}^{q}|\xi_j|$．依照平稳序列的谱表示定理(见[2])知

$$w_t=\int_{-\frac{1}{2}}^{\frac{1}{2}}e^{i2\pi t\lambda}\sum_{j=0}^{q}\zeta_j e^{-i2\pi\lambda j}dy(\lambda)\tag{1.6}$$

其中 $y(\lambda)$ 是正交增量过程，且 $E\{|dy(\lambda)|^2\}=\sigma^2 d\lambda$．令

$$a_t=\int_{-\frac{1}{2}}^{\frac{1}{2}}e^{i2\pi t\lambda}dy(\lambda)$$

易见 a_t 为白噪声序列，故由(1.6)式可知

$$w_t = \sum_{j=0}^{q} \zeta_j a_{t-j}$$

即 w_t 为白噪声序列的 q 阶滑动平均。

根据假定 $f(\lambda) \neq 0\left(-\frac{1}{2} \leqslant \lambda \leqslant \frac{1}{2}\right)$，即(1.5)式的根都不在单位圆上，那么，由于 $\xi_j(1 \leqslant j \leqslant q)$ 和 $\frac{1}{\xi_j}(1 \leqslant j \leqslant q)$ 是(1.5)式的 $2q$ 个根，且

$$|\xi_j|\left|\frac{1}{\xi_j}\right| = 1$$

所以这 $2q$ 个根中必有 q 个在单位圆外，另外 q 个在单位圆内，而且它们两两互为共轭倒数。我们将在单位圆外的 q 个根记做 $\eta_j(1 \leqslant j \leqslant q)$，于是仿前 $f(\lambda)$ 可以进一步表示成

$$f(\lambda) = \sigma_a^2 \left|\prod_{j=1}^{q}(1 - \eta_j^{-1} e^{-i2\pi\lambda})\right|^2 = \sigma_a^2 \left|\left(1 - \sum_{j=1}^{q} \theta_j e^{-i2\pi\lambda j}\right)\right|^2$$

其中 $\sigma_a^2 = |\gamma_q| \prod_{j=1}^{q} |\eta_j|$。依上述同样论证可得

$$w_t = a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q} \tag{1.7}$$

其中 a_t 为白噪声，$Ea_t^2 = \sigma_a^2$，且 $\theta(\omega) = 1 - \theta_1\omega - \cdots - \theta_q \omega^q$ 的根都在单位圆外，所以 w_t 是 MA$(0, q)$ 序列。

以上定理的逆定理是显然的，不再详细叙述了。

推论. 设 $w_t^{(l)} = \theta^{(l)}(B) a_t^{(l)} (l = 1, 2)$ 是两个相互独立的 $q_l(l = 1, 2)$ 阶 MA 序列，而且

$$\sigma_a^2(1)|\theta^{(1)}(e^{-i2\pi\lambda})|^2 + \sigma_a^2(2)|\theta^{(2)}(e^{-i2\pi\lambda})|^2 \neq 0$$

$$-\frac{1}{2} \leqslant \lambda \leqslant \frac{1}{2} \tag{1.8}$$

那么 $w_t = w_t^{(1)} + w_t^{(2)}$ 必是一个 $q(\leqslant \max(q_1, q_2))$ 阶的 MA 序列。

证明. 由于 $w_t^{(1)}$ 和 $w_t^{(2)}$ 是相互独立的，而且各自的协方差函

数 $\gamma_k(w^{(1)})$ 和 $\gamma_k(w^{(2)})$ 都是截尾的，所以 w_t 的自协方差函数

$$\gamma_k(w) = \gamma_k(w^{(1)}) + \gamma_k(w^{(2)})$$

也是截尾的，而且当 $k > \max(q_1, q_2)$ 时，$\gamma_k = 0$. 由定理 2 可知，w_t 是白噪声的 q 阶（$q \leqslant \max(q_1, q_2)$）滑动平均. 又易知，这时 w_t 的谱密度为

$$f(\lambda) = |\theta^{(1)}(e^{-i2\pi\lambda})|^2\sigma_a^2(1) + |\theta^{(2)}(e^{-i2\pi\lambda})|^2\sigma_a^2(2)$$

根据 (1.8) 式的假定，$f(\lambda)$ 处处不为零，所以由定理 2 进一步可知 w_t 是 MA$(0, q)$序列.

注意，只在 $q_1 = q_2$ 而且 $\theta_q^{(1)} = -\theta_q^{(2)}$ 时，才有 $q < \max(q_1, q_2)$，否则，$q = \max(q_1, q_2)$.

定理 3. 均值为零的平稳随机序列 w_t 为 AR $(p, 0)$ 序列的充分必要条件是，w_t 的偏相关函数 φ_{kk} 为 p 步截尾的（即 $\varphi_{pp} \neq 0$; $\varphi_{kk} = 0$, $k > p$）. φ_{kk} 由以下方程

$$\begin{pmatrix} 1 & \rho_1 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \cdots & \rho_{k-2} \\ \cdots & \cdots & \cdots & \cdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & 1 \end{pmatrix} \begin{pmatrix} \varphi_{k1} \\ \varphi_{k2} \\ \vdots \\ \varphi_{kk} \end{pmatrix} = \begin{pmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{pmatrix}, \quad k \geqslant 1 \tag{1.9}$$

的解唯一定义，其中 ρ_j 是 w_t 的自相关函数.

证明. 必要性. 若已知 w_t 为 AR$(p, 0)$ 序列，即

$$w_t - \varphi_1 w_{t-1} - \cdots - \varphi_p w_{t-p} = a_t \tag{1.10}$$

而且 $\varphi(\lambda) = 1 - \varphi_1\lambda - \cdots - \varphi_p\lambda^p$ 的根都在单位圆外. 则由第二章可知，φ_{kk} 是 p 步截尾的.

充分性. 若已知均值为零的平稳序列 w_t 的偏相关函数 $\varphi_{kk} = 0$，当 $k > p$，而 $\varphi_{pp} \neq 0$. 那么，根据第二章 (2.2.23) 式可知，当 $k \geqslant p$ 时

$$\varphi_{k+1,j} = \varphi_{kj} - \varphi_{k+1,k+1}\varphi_{k,k-(j-1)} = \varphi_{kj}, \quad 1 \leqslant j \leqslant k$$

因此，由数学归纳法容易推得，对于 $k \geqslant p+1$，

$$\varphi_{kj} = \begin{cases} \varphi_{pj}, & 1 \leqslant j \leqslant p \\ 0, & p+1 \leqslant j \leqslant k \end{cases}$$

再由第二章 (2.2.23) 式又知，当 $k \geqslant p$ 时，

$$0 = \varphi_{k+1,k+1} = \left(\rho_{k+1} - \sum_{j=1}^{k} \rho_{k+1-j}\varphi_{kj}\right)\left(1 - \sum_{j=1}^{k} \rho_j\varphi_{kj}\right)^{-1}$$

将前面关于 φ_{kj} 的两个表示式代人上式，即得

$$\rho_{k+1} - \sum_{j=1}^{p} \rho_{k+1-j}\varphi_{pj} = 0, \quad k \geqslant p$$

但 φ_{pj} 的定义 (1.9) 式直接表明，上式对 $0 \leqslant k < p$ 也成立，因而有

$$\rho_{k+1} - \sum_{j=1}^{p} \rho_{k+1-j}\varphi_{pj} = 0 \qquad k \geqslant 0$$

这个等式等价于 w_t 的自协方差函数 γ_k 满足方程

$$\gamma_k - \sum_{j=1}^{p} \varphi_{pj}\gamma_{k-j} = 0 \qquad k \geqslant 1$$

令

$$\varphi(B) \equiv 1 - \sum_{j=1}^{p} \varphi_{pj}B^j \tag{1.11}$$

那么前一式又可写做

$$\varphi(B)\gamma_k = 0, \qquad k \geqslant 1 \tag{1.12}$$

令

$$\varphi(B)w_t = a_t \tag{1.13}$$

那么由 (1.12) 式，当 $k \geqslant 1$ 时，

$$\begin{aligned} Ea_t a_{t+k} &= E\{\varphi(B)w_t\varphi(B)w_{t+k}\} \\ &= \varphi(B)\gamma_k - \varphi_{p1}\varphi(B)\gamma_{k+1} - \varphi_{p2}\varphi(B)\gamma_{k+2} \\ &\quad - \cdots - \varphi_{pp}\varphi(B)\gamma_{k+p} = 0 \end{aligned}$$

所以 a_t 是白噪声序列，而且对于 $k \geqslant 1$，

$$Ea_{t+k}w_t = \varphi(B)\gamma_k = 0 \tag{1.14}$$

现在，为证 w_t 是 AR(p, 0) 序列，只需证明 $\varphi(\omega)$ 的根都在单位圆外． 反设 $\varphi(\omega)$ 有不在单位圆外的根，不妨令 μ^{-1} 为这些根中的一个，$|\mu| \geqslant 1$，于是可记

$$\varphi(B) = (1 - \mu B)\varphi_1(B)$$

其中 $\varphi_1(B)$ 是 $p-1$ 次多项式. 令

$$z_t = \varphi_1(B) w_t$$

则

$$a_t = (1 - \mu B) z_t = z_t - \mu z_{t-1}$$

利用 (1.14) 式可知 $E a_t z_{t-1} = 0$, 因此由上式可得

$$E|z_t|^2 = E|a_t + \mu z_{t-1}|^2 = E a_t^2 + |\mu|^2 E|z_{t-1}|^2$$

由于 w_t 是平稳序列,所以 z_t 也是平稳序列,故$E|z_t|^2 = E|z_{t-1}|^2$. 又由于我们假定了 w_t 的偏相关函数存在, 所以 $(w_t, w_{t-1}, \cdots, w_{t-p})$ 不能线性相关, 即 a_t 的方差不可能为零. 于是由前一式得到

$$\begin{aligned} E|z_t|^2 &= E a_t^2 + |\mu|^2 E|z_{t-1}|^2 > |\mu|^2 E|z_{t-1}|^2 \\ &\geqslant E|z_{t-1}|^2 = E|z_t|^2 \end{aligned}$$

这就导致矛盾. 从而证明了 $\varphi(\omega)$ 的根只能都在单位圆外. 定理证毕.

§2. 样本自相关函数的渐近性质

一、无偏性与收敛性

设 w_t 是一正态 ARMA 序列, 容易看出样本自协方差函数

$$\hat{\gamma}_k = \frac{1}{N} \sum_{j=1}^{N-k} w_j w_{j+k}, \qquad k \geqslant 0 \tag{2.1}$$

是渐近无偏的,因为

$$E\hat{\gamma}_k = \frac{1}{N} \sum_{j=1}^{N-k} E w_j\, w_{j+k} = \frac{N-k}{N}\, \gamma_k = \gamma_k - \frac{k}{N} \gamma_k$$

不仅如此,对于更一般的

$$\hat{\gamma}_k^* = \frac{1}{N} \sum_{j=m_1+1}^{N-m_2} w_j w_{j+k}$$

$$E\hat{\gamma}_k^* = \gamma_k - \frac{m_1 + m_2}{N}\, \gamma_k \tag{2.2}$$

即 $\hat{\gamma}_k^*$ 也是渐近无偏的，其中 m_1, m_2 是与 N 无关的整数. 特别当 $m_1+m_2=0$ 时，$\hat{\gamma}_k^*$ 是 γ_k 的无偏估计.

根据 [2]（见 X § 7）可知，

$$\lim_{N\to\infty} \hat{\gamma}_k^* \overset{\text{a.s.}}{=} \gamma_k \tag{2.3}$$

以上收敛是指几乎处处收敛，因此 $\hat{\gamma}_k^*$ 也依概率收敛于 γ_k，即 $\hat{\gamma}_k^*$ 是 γ_k 的相容估计. 注意 (2.1) 式的 $\hat{\gamma}_k$ 是 $\hat{\gamma}_k^*$ 的一种特殊情况 ($m_1=0$, $m_2=k$)，所以 $\hat{\gamma}_k$ 也同样有这些性质. 由此便直接得知，w_t 的样本自相关函数

$$\hat{\rho}_k=\frac{\hat{\gamma}_k}{\hat{\gamma}_0} \quad \text{或} \quad \hat{\rho}_k^*=\frac{\hat{\gamma}_k^*}{\hat{\gamma}_0^*}$$

是 ρ_k 的相容估计，而且几乎处处收敛于 ρ_k. 但是关于 $\hat{\rho}_k$ 的渐近无偏性，要待以后才能证明.

二、高阶矩的渐近性质

关于估计误差 $\tilde{\gamma}_k=\hat{\gamma}_k-\gamma_k$, $\tilde{\rho}_k=\hat{\rho}_k-\rho_k$ 的各阶矩的性质，我们有以下定理.

定理 1. 设 w_t 为正态 ARMA 序列，那么

$$E\tilde{\gamma}_m\tilde{\gamma}_n=\frac{1}{N}\sum_{l=-\infty}^{\infty}(\gamma_{l-m}\gamma_{l-n}+\gamma_{l-m}\gamma_{l+n})+O\left(\frac{1}{N^2}\right) \tag{2.4}$$

证明. 由于 w_t 是正态平稳序列，所以它的有限维分布都是正态分布. 利用 [7] 中的公式

$$Ex_1x_2x_3x_4=Ex_1x_2Ex_3x_4+Ex_1x_3Ex_2x_4+Ex_1x_4Ex_2x_3 \tag{2.5}$$

其中 (x_1, x_2, x_3, x_4) 是任意均值为零的四维正态随机向量，以 w_i, w_{i+m}, w_j, w_{j+n} 代入上式可得

$$\begin{aligned}Ew_iw_{i+m}w_jw_{j+n}&=Ew_iw_{i+m}Ew_jw_{j+n}+Ew_iw_jEw_{i+m}w_{j+n}\\&\quad+Ew_iw_{j+n}Ew_jw_{i+m}\\&=\gamma_m\gamma_n+\gamma_{j-i}\gamma_{j-i+n-m}+\gamma_{j-i-m}\gamma_{j-i+n}\end{aligned}$$

由此式可以导出

$$
\begin{aligned}
E\tilde{\gamma}_n\tilde{\gamma}_m &= E\hat{\gamma}_n\hat{\gamma}_m - E\hat{\gamma}_n\gamma_m - E\gamma_n\hat{\gamma}_m + \gamma_n\gamma_m \\
&= E\hat{\gamma}_n\hat{\gamma}_m - \frac{N-n}{N}\gamma_n\gamma_m - \frac{N-m}{N}\gamma_n\gamma_m + \gamma_n\gamma_m \\
&= \frac{1}{N^2}\sum_{i=1}^{N-n}\sum_{j=1}^{N-m} E(w_i w_{i+n} w_j w_{j+m}) - \frac{N-(n+m)}{N}\gamma_n\gamma_m \\
&= \frac{1}{N^2}\sum_{i=1}^{N-n}\sum_{j=1}^{N-m} (\gamma_n\gamma_m + \gamma_{j-i}\gamma_{j-i+n-m} + \gamma_{j-i-m}\gamma_{j-i+n}) \\
&\quad - \frac{N-(n+m)}{N}\gamma_n\gamma_m \\
&= \frac{1}{N^2}\sum_{i=1}^{N-n}\sum_{j=1}^{N-m} (\gamma_{j-i}\gamma_{j-i+n-m} + \gamma_{j-i-m}\gamma_{j-i+n}) \\
&\quad + \frac{(N-n)(N-m)}{N^2}\gamma_m\gamma_n - \frac{N-(n+m)}{N}\gamma_m\gamma_n \\
&= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} (\gamma_{j-i}\gamma_{j-i+n-m} + \gamma_{j-i-m}\gamma_{j-i+n}) + O\left(\frac{1}{N^2}\right) \\
&= \frac{1}{N^2}\sum_{l=1-N}^{N-1} (N-|l|)(\gamma_l\gamma_{l+n-m} + \gamma_{l-m}\gamma_{l+n}) + O\left(\frac{1}{N^2}\right) \\
&= \frac{1}{N}\sum_{l=-\infty}^{\infty} (\gamma_l\gamma_{l+n-m} + \gamma_{l-m}\gamma_{l+n}) + O\left(\frac{1}{N^2}\right) \\
&\quad - \frac{1}{N^2}\sum_{l=1-N}^{N-1} |l|(\gamma_l\gamma_{l+m-n} + \gamma_{l-m}\gamma_{l+n})
\end{aligned}
$$

在式中改变和号的上下限时，须注意 $|\gamma_k| \leqslant c_1 e^{-c_2|k|}$（见第二章§2）。特别有

$$
\begin{aligned}
&\frac{1}{N^2}\left|\sum_{l=1-N}^{N-1} |l|(\gamma_l\gamma_{l+n-m} + \gamma_{l-m}\gamma_{l+n})\right| \\
&\leqslant \frac{\gamma_0}{N^2}\sum_{l=-\infty}^{\infty} |l|(|\gamma_l| + |\gamma_{l+n}|) = O\left(\frac{1}{N^2}\right)
\end{aligned}
$$

所以

$$
E\tilde{\gamma}_n\tilde{\gamma}_m = \frac{1}{N}\sum_{l=-\infty}^{\infty} (\gamma_l\gamma_{l+n-m} + \gamma_{l-m}\gamma_{l+n}) + O\left(\frac{1}{N^2}\right)
$$

$$= \frac{1}{N} \sum_{l=-\infty}^{\infty} (\gamma_{l-m}\gamma_{l-n} + \gamma_{l-m}\gamma_{l+n}) + O\left(\frac{1}{N^2}\right)$$

证明完毕.

此定理不难推广到 $\tilde{\gamma}_k^*$ 的情况，即对 $\tilde{\gamma}_k^*$ 而言仍满足 (2.4) 式. 又在 (2.4) 式中，取 $n = m$，即得到 $\hat{\gamma}_n$ 的方差估式

$$E\tilde{\gamma}_n^2 = \frac{1}{N} \sum_{l=-\infty}^{\infty} (\gamma_l^2 + \gamma_{l-n}\gamma_{l+n}) + O\left(\frac{1}{N^2}\right) \qquad (2.6)$$

推论. 当 w_t 为正态 ARMA 序列时，其样本自相关函数 $\hat{\rho}_k$ 满足

$$|E\tilde{\rho}_k| = O\left(\frac{1}{N}\right) \qquad (2.7)$$

证明. 不妨设 $\gamma_0 = 1$，于是

$$E\tilde{\rho}_k = E(\hat{\gamma}_k - \gamma_k\hat{\gamma}_0)\hat{\gamma}_0^{-1} = E(\hat{\gamma}_k - \gamma_k\hat{\gamma}_0)\left(1 + \frac{1 - \hat{\gamma}_0}{\hat{\gamma}_0}\right)$$

$$= E(\hat{\gamma}_k - \gamma_k\hat{\gamma}_0) - E(\hat{\gamma}_k - \gamma_k\hat{\gamma}_0)\tilde{\gamma}_0\hat{\gamma}_0^{-1}$$

$$= -\frac{k}{N}\gamma_k - E(\hat{\gamma}_k - \hat{\gamma}_k\hat{\gamma}_0 + \hat{\gamma}_k\hat{\gamma}_0 - \gamma_k\hat{\gamma}_0)\tilde{\gamma}_0\hat{\gamma}_0^{-1}$$

$$= E(\hat{\rho}_k\tilde{\gamma}_0 - \tilde{\gamma}_k)\tilde{\gamma}_0 + O\left(\frac{1}{N}\right)$$

由 $\hat{\gamma}_k$ 和 $\hat{\rho}_k$ 的定义可知

$$|\hat{\rho}_k| = \frac{\left|\sum_{l=1}^{N-k} w_{k+l}w_l\right|}{\left|\sum_{l=1}^{N} w_l^2\right|} \leqslant \frac{\left(\sum_{l=1}^{N-k} w_{k+l}^2 \sum_{l=1}^{N-k} w_l^2\right)^{\frac{1}{2}}}{\sum_{l=1}^{N} w_l^2} \leqslant 1$$

因此由 (2.4) 式有

$$|E\tilde{\rho}_k| \leqslant E|\hat{\rho}_k\tilde{\gamma}_0^2| + |E(\tilde{\gamma}_k\tilde{\gamma}_0)| + O\left(\frac{1}{N}\right)$$

$$\leqslant E\tilde{\gamma}_0^2 + |E(\tilde{\gamma}_k\tilde{\gamma}_0)| + O\left(\frac{1}{N}\right) = O\left(\frac{1}{N}\right)$$

此即 (2.7) 式. 这个推论还表明了，$\hat{\rho}_k$ 是 ρ_k 的渐近无偏估计.

定理 2. 设 w_t 为正态 ARMA 序列，那么对于任意整数 i_1, i_2, i_3, i_4 有以下渐近估式：

$$E\tilde{\gamma}_{i_1}\tilde{\gamma}_{i_2}\tilde{\gamma}_{i_3}=o\left(\frac{1}{N}\right)\qquad E(\tilde{\gamma}_{i_1}\tilde{\gamma}_{i_2}\tilde{\gamma}_{i_3}\tilde{\gamma}_{i_4})=o\left(\frac{1}{N}\right)\quad(2.8)$$

证明. 利用定理 1 可知，对任意整数 i，

$$E\tilde{\gamma}_i^2=O\left(\frac{1}{N}\right)$$

又因

$$|E\tilde{\gamma}_{i_1}\tilde{\gamma}_{i_2}\tilde{\gamma}_{i_3}|\leqslant\frac{1}{2}E(\tilde{\gamma}_{i_1}^2+\tilde{\gamma}_{i_2}^2)|\tilde{\gamma}_{i_3}|$$

$$\leqslant\frac{1}{2}\{(E\tilde{\gamma}_{i_1}^4)^{\frac{1}{2}}+(E\tilde{\gamma}_{i_2}^4)^{\frac{1}{2}}\}(E\tilde{\gamma}_{i_3}^2)^{\frac{1}{2}}$$

$$|E\tilde{\gamma}_{i_1}\tilde{\gamma}_{i_2}\tilde{\gamma}_{i_3}\tilde{\gamma}_{i_4}|\leqslant(E\tilde{\gamma}_{i_1}^2\tilde{\gamma}_{i_2}^2E\tilde{\gamma}_{i_3}^2\tilde{\gamma}_{i_4}^2)^{\frac{1}{2}}$$
$$\leqslant(E\tilde{\gamma}_{i_1}^4E\tilde{\gamma}_{i_2}^4E\tilde{\gamma}_{i_3}^4E\tilde{\gamma}_{i_4}^4)^{\frac{1}{4}}$$

所以为了证明 (2.8) 式，只须证明对任意整数 i，

$$E\tilde{\gamma}_i^4=o\left(\frac{1}{N}\right)\quad(2.9)$$

这是 (2.8) 式中后一式的一种特殊情况 ($i_1=i_2=i_3=i_4=i$).

首先有

$$E\tilde{\gamma}_m^3=E\hat{\gamma}_m^3-3\gamma_mE\hat{\gamma}_m^2+3\gamma_m^2E\hat{\gamma}_m-\gamma_m^3$$

$$=E\hat{\gamma}_m^3-3\gamma_mE(\tilde{\gamma}_m+\gamma_m)^2+3\gamma_m^2\frac{N-m}{N}\gamma_m-\gamma_m^3$$

$$=E\hat{\gamma}_m^3-\gamma_m^3+3\frac{m}{N}\gamma_m^3-3\gamma_mE\tilde{\gamma}_m^2$$

$$E\hat{\gamma}_m^3=\frac{1}{N^3}E\left(\sum_{k=1}^{N-m}w_kw_{k+m}\right)^3$$

$$=\frac{1}{N^3}\sum_{k}^{N-m}E(w_{k_1}w_{k_1+m}w_{k_2}w_{k_2+m}w_{k_3}w_{k_3+m})$$

上式中 $\sum\limits_{k}^{N-m}$ 是 $\sum\limits_{k_1,k_2,k_3=1}^{N-m}$ 的缩写记号. 仍然用 [7] 中求多维正态变量高阶矩的公式：

$$E w_{k_1} w_{k_1+m} w_{k_2} w_{k_2+m} w_{k_3} w_{k_3+m}$$

$$= \frac{1}{3!2^3} \sum_S E w_{i_1} w_{j_1} E w_{i_2} w_{j_2} E w_{i_3} w_{j_3}$$

其中求和范围 S 表示 $(i_1, j_1, i_2, j_2, i_3, j_3)$ 取遍 $(k_1, k_1+m, k_2, k_2+m, k_3, k_3+m)$ 的所有可能的排列. 我们将 S 分成三组: S_1 表示 $|i_1-j_1|=|i_2-j_2|=|i_3-j_3|=m$ 的各种排列; S_2 表示仅有一对 (i_s, j_s) 满足 $|i_s-j_s|=m$ 的各种排列; S_3 表示其余各种排列, 即 $|i_s-j_s|\neq m(s=1,2,3)$ 的各种排列. 于是

$$\begin{aligned}
E\hat{\gamma}_m^3 &= \frac{1}{3!2^3N^3}\sum_k^{N-m}\left(\sum_{S_1}+\sum_{S_2}+\sum_{S_3}\right)E w_{i_1}w_{j_1}E w_{i_2}w_{j_2}E w_{i_3}w_{j_3} \\
&= \frac{(N-m)^3}{N^3}\gamma_m^3+\frac{1}{3!2^3N^3}\sum_k^{N-m}3\cdot 3!2^3 \\
&\quad \cdot(\gamma_m\gamma_{k_2-k_3+m}\gamma_{k_2-k_3-m}+\gamma_m\gamma_{k_2-k_3}^2)+\frac{1}{3!2^3N^3} \\
&\quad \cdot\sum_k^{N-m}3!2^3(2\gamma_{k_1-k_2+m}\gamma_{k_2-k_3+m}\gamma_{k_1-k_3-m} \\
&\quad +6\gamma_{k_1-k_2}\gamma_{k_2-k_3+m}\gamma_{k_1-k_3})
\end{aligned} \tag{2.10}$$

而

$$\begin{aligned}
&\frac{1}{N^3}\left|\sum_k^{N-m}\gamma_{k_1-k_2+m}\gamma_{k_2-k_3+m}\gamma_{k_1-k_3-m}\right| \\
&= \frac{1}{N^3}\left|\sum_{k_1=1}^{N-m}\sum_{l=1-N+m}^{N-m-1}\sum_{k_2=\max(1,l+1)}^{\min(N-m+l,N-m)}\gamma_{k_1-k_2+m}\gamma_{l+m}\gamma_{k_1-k_2+l-m}\right| \\
&\leqslant \frac{\gamma_0}{N^3}\sum_{k_1,k_2=1}^{N-m}\sum_{l=-\infty}^{\infty}|\gamma_{k_1-k_2+m}\gamma_{l+m}|\leqslant\frac{\gamma_0}{N^2}\left(\sum_{j=-\infty}^{\infty}|\gamma_j|\right)^2=O\left(\frac{1}{N^2}\right)
\end{aligned}$$

同理

$$\frac{1}{N^3}\left|\sum_k^{N-m}\gamma_{k_1-k_2}\gamma_{k_2-k_3+m}\gamma_{k_1-k_3}\right|\leqslant O\left(\frac{1}{N^2}\right)$$

为了计算 (2.10) 式右边第二项的和值, 可利用定理 1 的证明方法估算出

$$\frac{1}{N^2}\sum_{k_1,k_3=1}^{N-m}(\gamma_{k_1-k_3+m}\gamma_{k_1-k_3-m}+\gamma_{k_1-k_3}^2)$$

$$=\frac{1}{N}\sum_{l=-\infty}^{\infty}(\gamma_l^2+\gamma_{l+m}\gamma_{l-m})+O\left(\frac{1}{N^2}\right)$$

将以上这些估式代入 $E\hat{\gamma}_m^3$ 可得

$$E\hat{\gamma}_m^3=\gamma_m^3-\frac{3m}{N}\gamma_m^3+\frac{3m^2}{N^2}\gamma_m^3-\frac{m^3}{N^3}\gamma_m^3+\frac{3(N-m)}{N^2}\gamma_m$$

$$\cdot\sum_{l=-\infty}^{\infty}(\gamma_l^2+\gamma_{l+m}\gamma_{l-m})+O\left(\frac{1}{N^2}\right)$$

$$=\gamma_m^3-\frac{3m}{N}\gamma_m^3+\frac{3\gamma_m}{N}\sum_{l=-\infty}^{\infty}(\gamma_l^2+\gamma_{l+m}\gamma_{l-m})+O\left(\frac{1}{N^2}\right)$$

再由 (2.6) 式有

$$E\tilde{\gamma}_m^2=\frac{1}{N}\sum_{l=-\infty}^{\infty}(\gamma_l^2+\gamma_{l+m}\gamma_{l-m})+O\left(\frac{1}{N^2}\right)$$

将这些估式代入 $E\tilde{\gamma}_m^3$ 便可得到

$$E\tilde{\gamma}_m^3=O\left(\frac{1}{N^2}\right)=o\left(\frac{1}{N}\right)\tag{2.11}$$

下面再证明 (2.9) 式. 利用 (2.11) 式可导出

$$E\tilde{\gamma}_m^4=E\hat{\gamma}_m^4-4\gamma_mE\hat{\gamma}_m^3+6\gamma_m^2E\hat{\gamma}_m^2-4\gamma_m^3E\hat{\gamma}_m+\gamma_m^4$$

$$=E\hat{\gamma}_m^4-4\gamma_mE(\tilde{\gamma}_m+\gamma_m)^3+6\gamma_m^2E(\tilde{\gamma}_m+\gamma_m)^2$$

$$-\frac{4(N-m)}{N}\gamma_m^4+\gamma_m^4$$

$$=E\hat{\gamma}_m^4-\gamma_m^4+\frac{4m}{N}\gamma_m^4-6\gamma_m^2E\tilde{\gamma}_m^2+O\left(\frac{1}{N^2}\right)$$

$$E\hat{\gamma}_m^4=\frac{1}{N^4}\sum_{k}^{N-m}Ew_{k_1}w_{k_1+m}w_{k_2}w_{k_2+m}w_{k_3}w_{k_3+m}w_{k_4}w_{k_4+m}$$

其中 $\sum\limits_{k}^{N-m}$ 是求和号 $\sum\limits_{k_1,k_2,k_3,k_4=1}^{N-m}$ 的缩写. 仍用[7]中求多维正态变量高阶矩的公式:

$$E w_{k_1} w_{k_1+m} w_{k_2} w_{k_2+m} w_{k_3} w_{k_3+m} w_{k_4} w_{k_4+m}$$
$$= \frac{1}{4!2^4} \sum_{S'} E w_{i_1} w_{j_1} w_{i_2} w_{j_2} w_{i_3} w_{j_3} w_{i_4} w_{j_4}$$

其中 S' 表示 $(i_1, j_1, i_2, j_2, i_3, j_3, i_4, j_4)$ 取遍 $(k_1, k_1+m, k_2, k_2+m, k_3, k_3+m, k_4, k_4+m)$ 的所有排列．仍分三组：S_1' 表示四对 (i_s, j_s) 同时满足 $|i_s - j_s| = m$ 的排列；S_2' 表示只有两对 (i_s, j_s) 满足 $|i_s - j_s| = m$ 的排列；S_3' 表示其余排列，即最多有一对 (i_s, j_s) 满足 $|i_s - j_s| = m$ 的排列．经过与前面类似的估算，可得到

$$E\hat{\gamma}_m^4 = \gamma_m^4 - \frac{4m}{N}\gamma_m^4 + \frac{6}{N}\gamma_m^2 \sum_{l=-\infty}^{\infty} (\gamma_l^2 + \gamma_{l+m}\gamma_{l-m}) + O\left(\frac{1}{N^2}\right)$$

再以 (2.6) 式和上式代入 $E\tilde{\gamma}_m^4$，便得到 (2.9) 式．定理证明完毕．

根据本定理的结果和证明方法，不难验证对 $\hat{\gamma}_k^*$ 也满足相应的 (2.8) 式．

推论 1. 在定理的条件下，对于任意的非零整数 j_1, j_2, j_3；

$$E(\tilde{\rho}_{j_1}\tilde{\rho}_{j_2}\tilde{\rho}_{j_3}) = o\left(\frac{1}{N}\right) \qquad E(\tilde{\rho}_{j_1}\tilde{\rho}_{j_2}\tilde{\rho}_{j_3}\tilde{\rho}_{j_4}) = o\left(\frac{1}{N}\right) \quad (2.12)$$

证明．由于

$$\tilde{\rho}_j = \frac{\hat{\gamma}_j\gamma_0 - \gamma_j\hat{\gamma}_0}{\gamma_0\hat{\gamma}_0} = \frac{\hat{\gamma}_j\gamma_0 - \hat{\gamma}_j\hat{\gamma}_0 + \hat{\gamma}_j\hat{\gamma}_0 - \gamma_j\hat{\gamma}_0}{\gamma_0\hat{\gamma}_0}$$
$$= \frac{\hat{\gamma}_0\tilde{\gamma}_j - \hat{\gamma}_j\tilde{\gamma}_0}{\gamma_0\hat{\gamma}_0} = \frac{\tilde{\gamma}_j}{\gamma_0} - \hat{\rho}_j\frac{\tilde{\gamma}_0}{\gamma_0}$$

并注意 $|\hat{\rho}_j| \leqslant 1$，应用定理 2，便可得到 (2.12) 式．

推论 2. 在定理的条件下，对于任意 m_1 个整数 $i_s (1 \leqslant s \leqslant m_1)$ 和 m_2 个非零整数 $j_\tau (1 \leqslant \tau \leqslant m_2)$，当 $m_1 + m_2 = 3$ 或 4 时

$$E\prod_{s=1}^{m_1} \tilde{\gamma}_{i_s} \prod_{\tau=1}^{m_2} \tilde{\rho}_{j_\tau} = o\left(\frac{1}{N}\right) \qquad (2.13)$$

证明．利用定理 2 的结论和推论 1 的方法便不难证明 (2.13) 式．

推论 3. 在定理的条件下，对任意非零整数 m, n，有

$$E\tilde{\rho}_m\tilde{\rho}_n = \frac{1}{N}\sum_{l=-\infty}^{\infty}\{\rho_{l-m}\rho_{l-n} + \rho_{l-m}\rho_{l+n} + 2\rho_m\rho_n\rho_l^2$$

$$- 2(\rho_n\rho_l\rho_{l+m} + \rho_m\rho_l\rho_{l+n})\} + o\left(\frac{1}{N}\right) \quad (2.14)$$

(2.14)式称为 Bartlett 公式.

证明. 不妨设 $\gamma_0 = 1$.

$$E\tilde{\rho}_m\tilde{\rho}_n = E\tilde{\rho}_m\tilde{\rho}_n\hat{\gamma}_0^2 + E\tilde{\rho}_m\tilde{\rho}_n(1 - \hat{\gamma}_0^2)$$

$$\begin{aligned} E\tilde{\rho}_m\tilde{\rho}_n\hat{\gamma}_0^2 &= E(\hat{\gamma}_m - \gamma_m\hat{\gamma}_0)(\hat{\gamma}_n - \gamma_n\hat{\gamma}_0) \\ &= E(\tilde{\gamma}_m - \gamma_m\tilde{\gamma}_0)(\tilde{\gamma}_n - \gamma_n\hat{\gamma}_0) \\ &= E\tilde{\gamma}_m\tilde{\gamma}_n + \gamma_m\gamma_n E\tilde{\gamma}_0^2 - \gamma_m E\tilde{\gamma}_n\tilde{\gamma}_0 - \gamma_n E\tilde{\gamma}_m\tilde{\gamma}_0 \end{aligned}$$

$$\begin{aligned} E\tilde{\rho}_m\tilde{\rho}_n(1 - \hat{\gamma}_0^2) &= E\tilde{\rho}_m\tilde{\rho}_n(1 - \hat{\gamma}_0)(1 + \hat{\gamma}_0) \\ &= -E\tilde{\rho}_m\tilde{\rho}_n\tilde{\gamma}_0 - E\tilde{\rho}_m\tilde{\rho}_n\tilde{\gamma}_0\hat{\gamma}_0 \\ &= E(\gamma_m\hat{\gamma}_0 - \hat{\gamma}_m)\tilde{\rho}_n\tilde{\gamma}_0 - E\tilde{\rho}_m\tilde{\rho}_n\tilde{\gamma}_0 \\ &= E(\gamma_m\tilde{\gamma}_0 - \tilde{\gamma}_m)\tilde{\rho}_n\tilde{\gamma}_0 - E\tilde{\rho}_m\tilde{\rho}_n\tilde{\gamma}_0 \end{aligned}$$

应用 (2.4) 和 (2.13) 式进一步计算以上两式便知，后一式的阶为 $o\left(\frac{1}{N}\right)$，而前一式恰好是 (2.14) 式的主项，从而证明了 (2.14) 式。

到此为止，我们给出了 $E\tilde{\gamma}_l^2$, $E\tilde{\rho}_l^2$ 等二阶矩的渐近公式，而且给出了更高阶矩的估算式 (2.13) 式. 这些结果在以后的几节中是很有用的.

三、样本自相关函数的渐近分布

为了讨论正态 ARMA 序列样本自相关函数的渐近分布，我们需要一些有关的预备概念、定理和引理.

设 $x_1, x_2\cdots$ 是随机序列，$Ex_t = \mu_t$, $E(x_t - \mu_t)^2 = \sigma_t^2$. 令

$$S_n = \sum_{k=1}^{n} x_k, \quad M_n = \sum_{k=1}^{n}\mu_k, \quad B_n^2 = \sum_{k=1}^{n}\sigma_k^2$$

如果当 $n\to\infty$ 时，$B_n^{-1}(S_n - M_n)$ 的分布渐近于正态 $\mathcal{N}(0,1)$分布，我们就说对此序列 x_t 中心极限定律成立.

中心极限定律成立是有条件的，在 x_t 为相互独立的情况下，

常用的判别准则是如下的定理，其证明可在一般的概率论教科书上找到(如文献[6]，[8])。

定理. 设 $x_1, x_2, \cdots$ 为相互独立的随机序列，而且 $Ex_t = 0$；令

$$\lambda_t^3 = E|x_t|^3, \qquad C_n^3 = \sum_{k=1}^{n} \lambda_k^3$$

若

$$\lim_{n\to\infty} B_n^{-1} C_n = 0$$

那么对 x_t 而言，中心极限定律成立。

此定理对一类更广的模型也成立，为今后引用方便，我们将它表成如下的引理(见文献[8])。

引理 1. 设对每个固定的 n, $x_1(n), x_2(n), \cdots, x_{\nu_n}(n)$ 是 ν_n 个相互独立的随机变量，$n = 1, 2, \cdots$；当 $n \to \infty$ 时，$\nu_n \to \infty$。又设

$$Ex_t(n) = 0, \quad Ex_t^2(n) = \sigma_t^2(n), \quad E|x_t(n)|^3 = \lambda_t^3(n)$$

$$B_n^2 = \sum_{k=1}^{\nu_n} \sigma_k^2(n), \qquad C_n^3 = \sum_{k=1}^{\nu_n} \lambda_k^3(n)$$

若

$$\lim_{n\to\infty} B_n^{-1} C_n = 0$$

那么对 $x_t(n)$ 而言，中心极限定律成立，即

$$B_n^{-1}\left(\sum_{t=1}^{\nu_n} x_t(n)\right)$$

的分布渐近于正态 $\mathcal{N}(0, 1)$ 分布。

若再加上条件

$$\lim_{n\to\infty} \nu_n^{-1} B_n^2 = \sigma^2$$

则

$$\nu_n^{-\frac{1}{2}}\left(\sum_{t=1}^{\nu_n} x_t(n)\right)$$

的分布渐近于正态 $\mathcal{N}(0, \sigma^2)$ 分布。

中心极限定律还可推广到更一般的情况. 设 $x_1, x_2, \cdots$ 是随机序列,若存在非负整数 m, 对于任意 $s-r>m$, $(x_1, x_2, \cdots x_r)$ 和 $(x_s, x_{s+1}, \cdots)$ 相互独立,我们就称这种序列为仅有 m 步依赖关系. 例如 m 阶正态滑动平均序列就具有这种性质. 我们要证明在加上某些条件之后，这样的序列也满足中心极限定律. 为此,首先假设 $Ex_t=0$, $Ex_t^2<\infty$; 并记

$$A_t = Ex_{t+m}^2 + 2\sum_{s=1}^{m} Ex_{t+m-s}x_{t+m}, \quad t\geqslant 1$$

显然,对于 $k>m$, 有

$$\begin{aligned}
&E(x_{t+1}+x_{t+2}+\cdots+x_{t+k})^2 \\
&= E(x_{t+1}+\cdots+x_{t+k-1})^2 + A_{t+k-m} \\
&= \cdots = E(x_{t+1}+\cdots+x_{t+m})^2 + \sum_{j=1}^{k-m} A_{t+j} \qquad (2.15)
\end{aligned}$$

引理 2. 设随机序列 $x_1, x_2, \cdots$ 仅有 m 步依赖关系,而且

$$E|x_t^3| \leqslant \lambda^3 < \infty, \qquad t\geqslant 1 \qquad (2.16)$$

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n} A_{t+k} = A \geqslant 0, \quad \text{对 } t \text{ 一致收敛} \qquad (2.17)$$

那么, $\frac{1}{\sqrt{n}}(x_1+x_2+\cdots+x_n)$ 的分布渐近于正态 $\mathcal{N}(0, A)$.

证明. 记

$$K=[n^{\frac{1}{5}}]=O(n^{\frac{1}{5}}), \qquad \nu=\left[\frac{n}{K}\right]=O(n^{\frac{4}{5}})$$

其中 $[\xi]$ 表示实数 ξ 的整数部分，则 $n=\nu K+r(0\leqslant r<K)$, $K^3=o(\nu)$. 再令

$$\begin{aligned}
S_n &= x_1+x_2+\cdots+x_n \\
&= (x_1+x_2+\cdots+x_K)+(x_{K+1}+x_{K+2}+\cdots+x_{2K}) \\
&\quad +\cdots+(x_{(\nu-1)K+1}+\cdots+x_{\nu K}) \\
&\quad + x_{\nu K+1}+\cdots+x_n
\end{aligned}$$

不妨设 $K>m$, 又记

$$x_{(i-1)K+1}+\cdots+x_{iK}$$
$$=(x_{(i-1)K+1}+\cdots+x_{iK-m})+x_{iK-m+1}+\cdots+x_{iK}$$
$$=U_i(n)+x_{iK-m+1}+\cdots+x_{iK}$$

于是

$$S_n=\sum_{i=1}^{\nu}U_i(n)+\sum_{i=1}^{\nu}(x_{iK-m+1}+\cdots+x_{iK})$$
$$+(x_{\nu K+1}+\cdots+x_n)=S'_n+T_n+V_n$$

其中 S'_n, T_n, V_n 分别代表上式中的三个求和项. 显然，依据引理的条件，对任意固定的 n, $U_1(n)$, $U_2(n)$, $\cdots$, $U_\nu(n)$ 是相互独立的，且 当 $n\to\infty$ 时，$\frac{1}{\sqrt{n}}S'_n=\frac{1}{\sqrt{n}}(\nu K)^{\frac{1}{2}}\left\{\nu^{-\frac{1}{2}}\sum_{i=1}^{\nu}K^{-\frac{1}{2}}U_i(n)\right\}$ 和 $\nu^{-\frac{1}{2}}\sum_{i=1}^{\nu}K^{-\frac{1}{2}}U_i(n)$ 这两个随机序列，或者有相同的极限分布，或者它们的分布都不收敛. 因为我们是讨论极限问题，只牵涉 n 充分大时的性质，所以不妨进一步设 $K>2m$. 这时由 (2.15) 式可知

$$\begin{aligned}E\{K^{-\frac{1}{2}}U_i(n)\}^2&=K^{-1}E(x_{(i-1)K+1}+\cdots+x_{iK-m})^2\\&=K^{-1}E(x_{(i-1)K+1}+\cdots+x_{(i-1)K+m})^2\\&\quad+K^{-1}\sum_{j=1}^{K-2m}A_{(i-1)K+j}\end{aligned}\tag{2.18}$$

又由 Hölder 不等式及条件 (2.16) 式可得

$$Ex_i^2\leqslant(E|x_i^2|^{\frac{3}{2}})^{\frac{2}{3}}=(E|x_i^3|)^{\frac{2}{3}}\leqslant\lambda^2\tag{2.19}$$

$$E|x_ix_j|\leqslant(Ex_i^2Ex_j^2)^{\frac{1}{2}}\leqslant\lambda^2\tag{2.20}$$

由此可见，(2.18) 式右边第一项将随 $K\to\infty$ 而一致趋于零. 根据 (2.17) 式，又知其第二项将随 $K\to\infty$ 而一致趋于 A，即

$$\lim_{n\to\infty}E(K^{-\frac{1}{2}}U_i(n))^2=A\text{，对于 } i \text{ 一致收敛}\tag{2.21}$$

其次，由 Hölder 不等式及 (2.16) 式知

$$E|x_ix_jx_r|\leqslant\{E|x_i^3|E|x_j^3|E|x_r^3|\}^{\frac{1}{3}}\leqslant\lambda^3$$

于是

$$E|K^{-\frac{1}{2}}U_i(n)|^3 = K^{-\frac{3}{2}}E\left|\sum_{j=1}^{K-m} x_{(i-1)K+j}\right|^3$$
$$\leqslant K^{-\frac{3}{2}}(K-m)^3\lambda^3 < K^{\frac{3}{2}}\lambda^3 = o(\nu^{\frac{3}{2}})$$
$$\sum_{i=1}^{\nu} E|K^{-\frac{1}{2}}U_i(n)|^3 = o(\nu^{\frac{3}{2}})$$
后一式左边相当在引理 1 中令 $x_i(n) = K^{-\frac{1}{2}}U_i(n)$ 所定出的 C_n^3. 而由 (2.21) 式知
$$\sum_{i=1}^{\nu} E[K^{-\frac{1}{2}}U_i(n)]^2 = O(\nu)$$
此式左边又相当在引理 1 中令 $x_i(n) = K^{-\frac{1}{2}}U_i(n)$ 所定出的 B_n^2. 由此可见,对于相互独立的 $K^{-\frac{1}{2}}U_i(n)(1 \leqslant i \leqslant \nu)$ 而言，引理 1 的条件 $B_n^{-1}C_n \to 0(n \to \infty)$ 是满足的,所以
$$\nu^{-\frac{1}{2}} \sum_{i=1}^{\nu} K^{-\frac{1}{2}}U_i(n)$$
的分布渐近于正态 $\mathscr{N}(0, A)$,从而 $n^{-\frac{1}{2}}S'_n$ 的极限分布也是 $\mathscr{N}(0, A)$. 再来考虑 $S_n = S'_n + T_n + V_n$ 中的另外两项. 由 (2.19) 和 (2.20) 式
$$n^{-1}ET_n^2 = n^{-1}\left[\sum_{i=1}^{\nu} E(x_{iK-m+1} + \cdots + x_{iK})^2\right]$$
$$\leqslant n^{-1}(\nu)m^2\lambda^2 = O(n^{-\frac{1}{3}})$$
$$n^{-1}EV_n^2 = n^{-1}E(x_{\nu K+1} + x_{\nu K+2} + \cdots + x_n)^2$$
$$\leqslant n^{-1}(n-\nu K)^2\lambda^2 \leqslant n^{-1}K^2\lambda^2 = O(n^{-\frac{1}{3}})$$
所以当 $n \to \infty$ 时,以上两式都趋于零,即 $n^{-\frac{1}{2}}T_n$ 和 $n^{-\frac{1}{2}}V_n$ 都均方收敛于零. 从而它们也依概率收敛于零. 因此根据[6](见第 6 章 § 3) 可知
$$n^{-\frac{1}{2}}S_n = n^{-\frac{1}{2}}S'_n + n^{-\frac{1}{2}}(T_n + V_n)$$
与 $n^{-\frac{1}{2}}S'_n$ 有相同的极限分布. 引理证毕.

推论. 设 x_t 是仅有 m 步依赖关系的平稳序列,而且 $Ex_t = 0$, $E|x_t^3|$ 有界，那么 $n^{-\frac{1}{2}}\sum_{t=1}^{n} x_t$ 的分布渐近于正态 $\mathscr{N}(0, A)$,其中

$$A = Ex_t^2 + 2\sum_{j=1}^{m} Ex_{t-j}x_t$$

证明. 由于 x_t 为平稳序列，而且仅有 m 步依赖关系，所以当 $n > m$ 时

$$0 < E\left(\sum_{t=1}^{n} x_t\right)^2 = E\left(\sum_{t=1}^{m} x_t\right)^2 + (n-m)A$$

由此可知 $A \geqslant 0$; 因为否则可取 n 足够大，使上式右边变为负数，而导致矛盾. 于是由引理 2 便得到此推论.

现在，我们用上述引理 1 和 2 来证明以下的定理.

定理 3. 设 w_t 为正态 ARMA(p, q) 序列，即为方程 (1, 1) 的正态平稳解. 为了简化记号，这里暂时令

$$\hat{\gamma}_k = \hat{\gamma}_k(n) = \frac{1}{n}\sum_{t=1}^{n} w_t w_{t+k}$$

那么，对于任何固定的正整数 K，当 $n \to \infty$ 时，随机变量 $\sqrt{n}(\hat{\gamma}_0 - \gamma_0)$, $\sqrt{n}(\hat{\gamma}_1 - \gamma_1)$, $\cdots$, $\sqrt{n}(\hat{\gamma}_K - \gamma_K)$ 的联合分布渐近于正态 $\mathcal{N}(0, G)$，其中 $\gamma_i = Ew_t w_{t+i}$, G 为 $K+1$ 阶方阵，其 i 行 j 列元素为

$$g_{ij} = \sum_{r=-\infty}^{\infty} (\gamma_{i+r}\gamma_{j+r} + \gamma_{i-r}\gamma_{j+r}), \quad i, j = 0, 1, \cdots, K \tag{2.22}$$

注意 g_{ij} 就是 (2.4) 式中 $nE\tilde{\gamma}_i\tilde{\gamma}_j$ 的主项.

证明. 先讨论 w_t 为纯滑动平均的情况，即

$$w_t = a_t - \sum_{j=1}^{q} \theta_j a_{t-j} = -\sum_{j=0}^{q} \theta_j a_{t-j}, \quad \theta_0 = -1$$

任给 $K+1$ 个实数 $u_0, u_1, \cdots, u_K$，令

$$x_t = u_0(w_t^2 - \gamma_0) + u_1(w_t w_{t+1} - \gamma_1) + \cdots + u_K(w_t w_{t+K} - \gamma_K)$$

显然，x_t 为仅有 $m = K + q$ 步依赖关系的平稳序列，$Ex_t = 0$, $E|x_t|^3$ 存在且与 t 无关. 依引理 2 的推论知

$$U_n = n^{-\frac{1}{2}} \sum_{t=1}^{n} x_t = u_0[n^{\frac{1}{2}}(\hat{\gamma}_0 - \gamma_0)] + \cdots + u_K[n^{\frac{1}{2}}(\hat{\gamma}_K - \gamma_K)]$$

的极限分布为正态 $\mathcal{N}(0, A)$，其中

$$A = Ex_t^2 + 2\sum_{r=1}^{m} Ex_t x_{t+r} = \sum_{r=-m}^{m} Ex_t x_{t+r}$$

$$= \sum_{i,j=0}^{K} u_i u_j g_{ij} \tag{2.23}$$

而

$$g_{ij} = \sum_{r=-m}^{m} E[(w_t w_{t+i} - \gamma_i)(w_{t+r} w_{t+r+j} - \gamma_j)]$$

再一次应用 (2.5) 式并注意 $\gamma_k = 0(|k| > m)$ 可得

$$g_{ij} = \sum_{r=-m}^{m} [Ew_t w_{t+i} w_{t+r} w_{t+r+j} - \gamma_i \gamma_j]$$

$$= \sum_{r=-m}^{m} (\gamma_i \gamma_j + \gamma_r \gamma_{j-i+r} + \gamma_{j+r} \gamma_{i-r} - \gamma_i \gamma_j)$$

$$= \sum_{r=-\infty}^{\infty} (\gamma_r \gamma_{j-i+r} + \gamma_{j+r} \gamma_{i-r})$$

$$= \sum_{r=-\infty}^{\infty} (\gamma_{i+r} \gamma_{j+r} + \gamma_{j+r} \gamma_{i-r})$$

若以 G 表示由 $g_{ij}(i, j = 0, 1, \cdots, K)$ 所构成的 $K+1$ 阶方阵．由于 $0 \leqslant A = \boldsymbol{u}^\tau G \boldsymbol{u}$ 和 $\boldsymbol{u} = (u_0, u_1, \cdots, u_K)^\tau$ 的任意性，可知 G 为非负定阵．又由于 U_n 的分布收敛于 $\mathcal{N}(0, A)$，所以其特征函数 $\varphi_n(v) = Ee^{ivU_n}$ 收敛于 $e^{-\frac{A}{2}v}$．特别取 $v = 1$ 知，Ee^{iU_n} 收敛于 $e^{-\frac{A}{2}}$，即对任意的 $K+1$ 维向量 $\boldsymbol{u}$，$Ee^{i\sum_{j=0}^{K} u_j \sqrt{n}(\hat{\gamma}_j - \gamma_j)}$ 收敛于 $e^{-\frac{1}{2}\boldsymbol{u}^\tau G \boldsymbol{u}}$．这表明，随机向量 $(n^{\frac{1}{2}}(\hat{\gamma}_0 - \gamma_0), n^{\frac{1}{2}}(\hat{\gamma}_1 - \gamma_1), \cdots, n^{\frac{1}{2}}(\hat{\gamma}_K - \gamma_K))$ 的特征函数收敛于 $K+1$ 维正态分布 $\mathcal{N}(0, G)$ 的特征函数，从而证明了随机向量的极限分布为 $\mathcal{N}(0, G)$．

下面讨论 w_t 为一般的正态 ARMA(p, q) 序列的情况．先将 w_t 写成逆转形式

$$w_t = \varphi^{-1}(B)\theta(B)a_t = \psi(B)a_t = \sum_{j=0}^{\infty}\psi_j a_{t-j}$$

在第二章中已经指出过，$|\psi_j| \leqslant c_1 e^{-c_2 j}$（$c_1$，$c_2$ 表示两个正常数）。给定正整数 Q，暂记

$$x_t = \sum_{j=0}^{Q}\psi_j a_{t-j}, \qquad y_t = \sum_{j=Q+1}^{\infty}\psi_j a_{t-j}, \qquad w_t = x_t + y_t$$

于是 x_t，y_t 的自协方差函数及它们的互协方差函数分别为

$$\gamma_{-k}(x) = \gamma_k(x) = \begin{cases}\sum_{j=0}^{Q-k}\psi_j\psi_{j+k}\sigma_a^2, & 0 \leqslant k \leqslant Q\\ 0, & k > Q\end{cases}$$

$$\gamma_{-k}(y) = \gamma_k(y) = \sum_{j=Q+1}^{\infty}\psi_j\psi_{j+k}\sigma_a^2, \quad k \geqslant 0$$

$$\gamma_k(xy) = \begin{cases}\sum_{j=\max(0,Q-k+1)}^{Q}\psi_j\psi_{j+k}\sigma_a^2, & k > 0\\ 0, & k \leqslant 0\end{cases}$$

由此易知，$\gamma_0(x) \leqslant \gamma_0$，$\gamma_0(y) \leqslant \gamma_0$，$|\gamma_l(xy)|^2 \leqslant \gamma_0(x)\gamma_0(y) \leqslant \gamma_0^2$，且对 $k > 0$

$$|\gamma_k(y)| \leqslant \sum_{j=Q+1}^{\infty}|\psi_j||\psi_{j+k}|\sigma_a^2 \leqslant \sigma_a^2 c_1^2 \sum_{j=Q+1}^{\infty} e^{-c_2(2j+k)}$$
$$= \sigma_a^2 c_1^2 e^{-c_2 k} e^{-2c_2 Q} e^{-2c_2}/(1 - e^{-2c_2}) \equiv M_1 e^{-c_2 k} e^{-2c_2 Q}$$

类似地，对 $k > 0$

$$|\gamma_k(xy)| \leqslant M_2 e^{c_2 k} e^{-2c_2 Q}$$

现在，把 $\tilde{\gamma}_k$ 分成如下两部分：

$$\tilde{\gamma}_k = \hat{\gamma}_k - \gamma_k = \frac{1}{n}\sum_{t=1}^{n}(w_t w_{t+k} - \gamma_k)$$
$$= \frac{1}{n}\sum_{t=1}^{n}[(x_t + y_t)(x_{t+k} + y_{t+k}) - \gamma_k)$$
$$= \frac{1}{n}\sum_{t=1}^{n}\{x_t x_{t+k} - \gamma_k(x)\}$$
$$+ \frac{1}{n}\sum_{t=1}^{n}(x_t y_{t+k} + x_{t+k}y_t + y_t y_{t+k} - \gamma_k + \gamma_k(x))$$
$$\equiv \tilde{\gamma}_k(x) + \varepsilon_k(Q) \tag{2.24}$$

其中

$$\tilde{\gamma}_k(x) = \hat{\gamma}_k(x) - \gamma_k(x) = \frac{1}{n}\sum_{t=1}^{n}(x_t x_{t+k} - \gamma_k(x))$$

$$\varepsilon_k(Q) = \frac{1}{n}\sum_{t=1}^{n}[x_t y_{t+k} + x_{t+k}y_t + y_t y_{t+k} - \gamma_k + \gamma_k(x)]$$

$$= \frac{1}{n}\sum_{t=1}^{n}\{[x_t y_{t+k} - \gamma_k(xy)]$$

$$+ [x_{t+k}y_t - \gamma_{-k}(xy)] + [y_t y_{t+k} - \gamma_k(y)]\}$$

用(2.5)式可计算出

$$\frac{1}{n^2}\, E\left\{\sum_{t=1}^{n}(x_t y_{t+k} - \gamma_k(xy))\right\}^2$$

$$= \frac{1}{n^2}\sum_{t,s=1}^{n}(E x_t y_{t+k} x_s y_{s+k} - \gamma_k(xy)^2)$$

$$= \frac{1}{n^2}\sum_{t,s=1}^{n}(\gamma_{t-s}(x)\gamma_{t-s}(y) + \gamma_{s-t+k}(xy)\gamma_{t-s+k}(xy))$$

$$= \frac{1}{n}\sum_{l=-Q}^{Q}\gamma_l(x)\,\gamma_l(y) + \frac{1}{n}\sum_{l=1-k}^{k-1}\gamma_{k+l}(xy)\,\gamma_{k-l}(xy)$$

$$+ O\left(\frac{1}{n^2}\right)$$

对上式右边前两项求和值，可分别作以下估算：

$$\left|\frac{1}{n}\sum_{l=-Q}^{Q}\gamma_l(x)\gamma_l(y)\right| \leqslant \frac{1}{n}\gamma_0\sum_{l=-\infty}^{\infty}|\gamma_l(y)|$$

$$\leqslant \frac{1}{n}\gamma_0 M_1 e^{-2c_2 Q}\sum_{l=-\infty}^{\infty}e^{-c_2|l|} \equiv \frac{1}{n}M_3 e^{-2c_2 Q}$$

$$\left|\frac{1}{n}\sum_{l=1-k}^{k-1}\gamma_{k+l}(xy)\gamma_{k-l}(xy)\right| \leqslant \frac{1}{n}\gamma_0\sum_{l=1-k}^{k-1}|\gamma_{k+l}(xy)|$$

$$\leqslant \frac{1}{n}\gamma_0 M_2 e^{-2c_2 Q}\sum_{l=1-k}^{k-1}e^{c_2(k+l)} \equiv \frac{1}{n}M_4 e^{-2c_2 Q}$$

其中 M_3 和 M_4 是最多只与 k 有关的正常数. 按同样的办法可估算 $\varepsilon_k(Q)$ 表达式中其余两项的方差，于是联合可得

$$E(\varepsilon_k(Q))^2 \leqslant 3\,\frac{1}{n^2}\left[E\left\{\sum_{t=1}^{n}(x_t y_{t+k}-\gamma_k(xy))\right\}^2\right.$$

$$+E\left\{\sum_{t=1}^{n}(x_{t+k}y_t-\gamma_{-k}(xy))\right\}^2$$

$$\left.+E\left\{\sum_{t=1}^{n}(y_t y_{t+k}-\gamma_k(y))\right\}^2\right]$$

$$\leqslant \frac{1}{n}Me^{-2c_2Q}+O\left(\frac{1}{n^2}\right) \tag{2.25}$$

其中M是一最多只与 k 有关(而与 Q 和 n 无关)的正常数.

现在，我们利用上面的结果来证明，$\sqrt{n}\,\tilde{\gamma}_0$, $\sqrt{n}\,\tilde{\gamma}_1$, $\cdots$, $\sqrt{n}\,\tilde{\gamma}_K$ 的联合分布渐近于正态 $\mathscr{N}(0, G)$. 依据与本定理的前一部分论证相同的理由，我们只须证明，对任给的 $K+1$ 维向量 $\boldsymbol{u}=(u_0, u_1, \cdots, u_K)^\tau$, $(K<Q)$, 随机变量

$$U_n=\sum_{k=0}^{K}u_k\sqrt{n}\,\tilde{\gamma}_k$$

的极限分布为 $\mathscr{N}(0, \boldsymbol{u}^\tau G\boldsymbol{u})$. 由(2.24)式，可以把 U_n 分为两项之和:

$$U_n=\sum_{k=0}^{K}u_k\sqrt{n}\,\tilde{\gamma}_k(x)+\sum_{k=0}^{K}u_k\sqrt{n}\,\varepsilon_k(Q)$$
$$=U_n(Q)+V_n(Q)$$

又由(2.25)式知

$$E(V_n(Q))^2=nE\left(\sum_{k=0}^{K}u_k\varepsilon_k(Q)\right)^2\leqslant nK\sum_{k=0}^{K}u_k^2E(\varepsilon_k(Q))^2$$

$$\leqslant nK\left(\sum_{k=0}^{K}u_k^2\right)\max_{0\leqslant k\leqslant K}E(\varepsilon_k(Q))^2\leqslant ce^{-2c_2Q}+O\left(\frac{1}{n}\right) \tag{2.26}$$

其中 c 是一个与 Q 和 n 都无关的正常数.

设 ξ 为任一实数，ε 为任一正实数，则一方面有

$$
\begin{aligned}
P\{U_n<\xi\}&=P\{U_n(Q)+V_n(Q)<\xi\}\\
&=P\{U_n(Q)<\xi-V_n(Q),\ |V_n(Q)|<\varepsilon\}\\
&\quad+P\{U_n(Q)<\xi-V_n(Q),\ |V_n(Q)|\geqslant\varepsilon\}\\
&\leqslant P\{U_n(Q)<\xi+\varepsilon\}+P\{|V_n(Q)|\geqslant\varepsilon\}
\end{aligned}
$$

另一方面有

$$
\begin{aligned}
P\{U_n(Q)<\xi-\varepsilon\}&=P\{U_n-V_n(Q)<\xi-\varepsilon\}\\
&=P\{U_n<\xi-\varepsilon+V_n(Q),\ |V_n(Q)|<\varepsilon\}\\
&\quad+P\{U_n<\xi-\varepsilon+V_n(Q),|V_n(Q)|\geqslant\varepsilon\}\\
&\leqslant P\{U_n<\xi\}+P\{|V_n(Q)|\geqslant\varepsilon\}
\end{aligned}
$$

合并两式即得

$$
\begin{aligned}
P\{U_n(Q)<\xi-\varepsilon\}-P\{|V_n(Q)|\geqslant\varepsilon\}&\leqslant P\{U_n<\xi\}\\
&\leqslant P\{U_n(Q)<\xi+\varepsilon\}+P\{|V_n(Q)|\geqslant\varepsilon\} \qquad (2.27)
\end{aligned}
$$

注意 $x_t=\sum_{j=0}^{Q}\phi_j a_{t-j}$ 是 Q 阶纯滑动平均序列，因此由定理证明的前一部分，$U_n(Q)=\sum_{k=0}^{K}u_k\sqrt{n}\,\tilde{\gamma}_k(x)$ 的极限分布是 $\mathcal{N}(0,\ \boldsymbol{u}^{\tau}\cdot G(Q)\boldsymbol{u})$，其中 $G(Q)$ 的 i 行 j 列元为

$$
g_{ij}(Q)=\sum_{r=-\infty}^{\infty}(\gamma_{i+r}(x)\gamma_{j+r}(x)+\gamma_{i-r}(x)\gamma_{j+r}(x))
$$

$$
i,j=0,\ 1,\ \cdots,\ K
$$

也就是说，当 $n\to\infty$ 时，

$$
\begin{aligned}
P\{U_n(Q)<\xi\}\to\Phi_Q(\xi)&=\frac{1}{\sqrt{2\pi\boldsymbol{u}^{\tau}G(Q)\boldsymbol{u}}}\\
&\quad\times\int_{-\infty}^{\xi}e^{-\frac{1}{2}(\boldsymbol{u}^{\tau}G(Q)\boldsymbol{u})^{-1}\eta^2}d\eta
\end{aligned}
$$

又由 (2.26) 式知

$$
\begin{aligned}
P\{|V_n(Q)|\geqslant\varepsilon\}&\leqslant\frac{1}{\varepsilon^2}E(V_n(Q))^2\\
&\leqslant\frac{c}{\varepsilon^2}e^{-k_1Q}+O\left(\frac{1}{\varepsilon^2 n}\right)
\end{aligned}
$$

于是，令 $n \to \infty$，对 (2.27) 式两边分别取下极限和上极限，就得到以下的不等式：

$$\Phi_Q(\xi - \varepsilon) - \frac{c}{\varepsilon^2} e^{-2x_2 Q} \leqslant \varliminf_{n \to \infty} P\{U_n < \xi\} \leqslant \varlimsup_{n \to \infty} P\{U_n < \xi\}$$

$$\leqslant \Phi_Q(\xi + \varepsilon) + \frac{c}{\varepsilon^2} e^{-2x_2 Q}$$

其中 Q 可以是任意大的正整数，ε 可以是任意小的正实数。最后，在上式中先令 $Q \to \infty$，再令 $\varepsilon \to 0$，即得

$$\lim_{n \to \infty} P\{U_n < \xi\} = \Phi(\xi)$$

其中 $\Phi(\xi)$ 是正态 $\mathscr{N}(0, u^\tau G u)$ 的分布函数：

$$\Phi(\xi) = \frac{1}{\sqrt{2\pi u^\tau G u}} \int_{-\infty}^{\xi} e^{-\frac{1}{2}(u^\tau G u)^{-1}\eta^2} d\eta$$

而 G 的第 i 行 j 列元为

$$g_{ij} = \lim_{Q \to \infty} g_{ij}(Q) = \sum_{r=-\infty}^{\infty} (\gamma_{i+r}\gamma_{j+r} + \gamma_{i-r}\gamma_{j+r})$$

$$i, j = 0, 1, \cdots, K.$$

至此全部证明了定理.

这个定理的结论可以推广到非正态 ARMA 序列的情形，只须假定序列 w_t 模型中的白噪声 a_t 有有限的四阶矩：$Ea_t^4 < \infty$。这里我们不详细讨论了.

推论. 设 w_t 为正态平稳白噪声序列，则随机向量 $(\sqrt{n}(\hat{\gamma}_0 - \gamma_0), \sqrt{n}(\hat{\gamma}_1 - \gamma_1), \cdots, \sqrt{n}(\hat{\gamma}_K - \gamma_K))$ 的极限分布为 $\mathscr{N}(0, G)$，其中

$$G = \gamma_0^2 \begin{pmatrix} 2 & 0 \\ 0 & I_K \end{pmatrix}, \qquad (I_K \text{ 表 } K \text{ 阶单位阵})$$

它由定理 1 直接推得. 由此可知，对于白噪声，$\sqrt{n}(\hat{\gamma}_k - \gamma_k)$ $(k = 0, 1, 2, \cdots)$ 为相互渐近独立.

为了将定理 3 中关于样本协方差函数的结果应用于样本相关函数，我们还需以下引理.

引理 3. 设 $\boldsymbol{x}_n$ 是 r 维随机向量列，当 $n\to\infty$ 时，其分布 $F_n(\boldsymbol{x})$ 收敛于分布 $F(\boldsymbol{x})$；又设随机变量列 y_n 依概率收敛于正常数 b。那么，随机向量列 $\frac{\boldsymbol{x}_n}{y_n}$ 的分布收敛于分布 $F(b\boldsymbol{x})$。

证明. 设 ε 是任一正实数，又设 $\boldsymbol{x}=(x_1, x_2, \cdots, x_r)^{\tau}$ 是任一 r 维常向量. 暂用 $\{\boldsymbol{x}_n<\boldsymbol{x}\}$ 表示集合 $\{x_{ni}<x_i, 1\leqslant i\leqslant r\}$，其中 x_{ni} 表示 $\boldsymbol{x}_n$ 的分量，于是

$$P\left\{\frac{\boldsymbol{x}_n}{y_n}<\boldsymbol{x},\ |y_n-b|<\varepsilon\right\}\leqslant P\left\{\frac{\boldsymbol{x}_n}{y_n}<\boldsymbol{x}\right\}$$
$$\leqslant P\left\{\frac{\boldsymbol{x}_n}{y_n}<\boldsymbol{x},\ |y_n-b|<\varepsilon\right\}+P\{|y_n-b|\geqslant\varepsilon\} \tag{2.28}$$

令

$$\Pi=\begin{pmatrix}\operatorname{sgn}x_1 & & & 0\\ & \operatorname{sgn}x_2 & & \\ & & \ddots & \\ 0 & & & \operatorname{sgn}x_r\end{pmatrix}$$

其中 $\operatorname{sgn}x_i$ 依 x_i 为正、负或零而分别等于 $+1$, -1 或 0，又用 I 表 r 阶单位阵，则易知有不等式

$$\begin{aligned}F_n((bI+\varepsilon\Pi)\boldsymbol{x})&=P\{\boldsymbol{x}_n<(bI+\varepsilon\Pi)\boldsymbol{x}\}\\&\geqslant P\{\boldsymbol{x}_n<(bI+\varepsilon\Pi)\boldsymbol{x},\ b-\varepsilon<y_n<b+\varepsilon\}\\&\geqslant P\{\boldsymbol{x}_n<y_n\boldsymbol{x},\ |y_n-b|<\varepsilon\}\\&\geqslant P\{\boldsymbol{x}_n<(bI-\varepsilon\Pi)\boldsymbol{x},\ |y_n-b|<\varepsilon\}\\&\geqslant F_n((bI-\varepsilon\Pi)\boldsymbol{x})-P\{|y_n-b|\geqslant\varepsilon\}\end{aligned}$$

以此代入 (2.28) 式得

$$F_n((bI-\varepsilon\Pi)\boldsymbol{x})-P\{|y_n-b|\geqslant\varepsilon\}\leqslant P\left\{\frac{\boldsymbol{x}_n}{y_n}<\boldsymbol{x}\right\}$$
$$\leqslant F_n((bI+\varepsilon\Pi)\boldsymbol{x})+P\{|y_n-b|\geqslant\varepsilon\}$$

注意 y_n 依概率收敛于 b，$F_n(\boldsymbol{x})$ 收敛于 $F(\boldsymbol{x})$，由此有

$$F((bI-\varepsilon\Pi)\boldsymbol{x})\leqslant\varliminf_{n\to\infty}P\left\{\frac{\boldsymbol{x}_n}{y_n}<\boldsymbol{x}\right\}\leqslant\varlimsup_{n\to\infty}P\left\{\frac{\boldsymbol{x}_n}{y_n}<\boldsymbol{x}\right\}$$
$$\leqslant F((bI+\varepsilon\Pi)\boldsymbol{x})$$

再令 $\varepsilon \to 0$ 即得

$$F(bx) = \lim_{n\to\infty} P\left\{\frac{x_n}{y_n} < x\right\}$$

引理证毕.

推论. 若随机向量列 x_n 的分布收敛于正态分布 $\mathcal{N}(0, R)$，随机变量列 y_n 依概率收敛于常数 b，则 $\frac{x_n}{y_n}$ 的分布收敛于正态分布 $\mathcal{N}\left(0, \frac{1}{b^2}R\right)$.

定理 4. 设 w_t 为正态 ARMA 序列，ρ_k 为其自相关函数，$\hat{\rho}_k = \hat{\gamma}_k/\hat{\gamma}_0$ 为其样本自相关函数，$\tilde{\rho}_k = \hat{\rho}_k - \rho_k$. 那么$\sqrt{n}\,\tilde{\rho}_1, \sqrt{n}\,\tilde{\rho}_2, \cdots, \sqrt{n}\,\tilde{\rho}_K$ 的联合分布渐近于K 维正态分布 $\mathcal{N}(0, B)$，其中B 的 i 行 j 列元素为

$$\begin{aligned} b_{ij} = \sum_{r=-\infty}^{\infty} [&\rho_{i+r}\rho_{j+r} + \rho_{i-r}\rho_{j+r} + 2\rho_i\rho_j\rho_r^2 \\ &- 2\rho_r(\rho_i\rho_{j+r} + \rho_j\rho_{i+r})] \quad i, j = 1, 2, \cdots, K \end{aligned} \tag{2.29}$$

即是 (2.14) 式中 $nE\tilde{\rho}_i\tilde{\rho}_j$ 的主项. 特别当 w_t 为正态白噪声序列时，$B = I$.

证明. 令

$$\xi_n^{\tau} = (\sqrt{n}\,\tilde{\rho}_1, \cdots, \sqrt{n}\,\tilde{\rho}_K)$$

$$\zeta_n^{\tau} = (\sqrt{n}\,\tilde{\gamma}_0, \sqrt{n}\,\tilde{\gamma}_1, \cdots, \sqrt{n}\,\tilde{\gamma}_K)$$

则由定理 3 知，ζ_n 的分布收敛于正态 $\mathcal{N}(0, G)$. 又由定理 1 知，对任给的正数 ε，当 $n \to \infty$ 时，

$$P\{|\hat{\gamma}_0 - \gamma_0| \geqslant \varepsilon\} \leqslant \frac{1}{\varepsilon^2} E\tilde{\gamma}_0^2 = \frac{2}{\varepsilon^2 n}\sum_{l=-\infty}^{\infty}\gamma_l^2 + O\left(\frac{1}{\varepsilon^2 n^2}\right) \to 0$$

即 $\hat{\gamma}_0$ 依概率收敛于 γ_0. 另一方面，由于

$$\tilde{\rho}_k = \hat{\rho}_k - \rho_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} - \frac{\gamma_k}{\hat{\gamma}_0} + \frac{\gamma_k}{\hat{\gamma}_0} - \frac{\gamma_k}{\gamma_0} = \frac{\tilde{\gamma}_k}{\hat{\gamma}_0} - \rho_k\frac{\tilde{\gamma}_0}{\hat{\gamma}_0}$$

所以

$$\xi_n = H\zeta_n/\hat{\gamma}_0$$

其中H是如下的 $K\times(K+1)$ 阶矩阵:

$$H=\begin{pmatrix} -\rho_1 & 1 & 0 & \cdots & 0 \\ -\rho_2 & 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ -\rho_k & 0 & \cdots & \cdots & 1 \end{pmatrix}_{K\times(K+1)}$$

因为$H\zeta_n$的分布收敛于正态 $\mathcal{N}(0, HGH^\tau)$，再利用引理 3 的推论，即知ξ_n的分布收敛于正态分布 $N\left(0, \dfrac{1}{\gamma_0^2}HGH^\tau\right)$，因此，$B=\dfrac{1}{\gamma_0^2}HGH^\tau$. 由此和 (2.22) 式不难算出$B$的具体表达式 (2.29).

当 w_t为正态白噪声时，利用 (2.29) 式，容易验证 $B=I$. 定理证毕.

§3. 正态 ARMA 序列的分布密度

正态 ARMA 序列的分布密度函数被模型参数 $\varphi_1, \varphi_2, \cdots, \varphi_p$; $\theta_1, \theta_2, \cdots, \theta_q$ 和 σ_a^2 唯一决定. 我们仍采用第二章里引入的参数向量记号:

$$\boldsymbol{\varphi}^\tau=(\varphi_1, \varphi_2, \cdots, \varphi_p), \quad \boldsymbol{\theta}^\tau=(\theta_1, \theta_2, \cdots, \theta_q)$$

$$\boldsymbol{\beta}^\tau=(\boldsymbol{\varphi}^\tau, \boldsymbol{\theta}^\tau), \qquad \boldsymbol{\alpha}^\tau=(\boldsymbol{\beta}^\tau, \sigma_a^2)$$

依第四章所介绍的分布密度函数的定义可知，$\boldsymbol{w}_n=(w_1, w_2, \cdots, w_n)^\tau$ 的概率分布密度为

$$p(\boldsymbol{w}_n|\boldsymbol{\beta}, \sigma_a^2)=(2\pi\sigma_a^2)^{-\frac{n}{2}}|M_n|^{\frac{1}{2}}\exp\left\{-\frac{S_n(\boldsymbol{\beta})}{2\sigma_a^2}\right\} \tag{3.1}$$

其中

$$S_n(\boldsymbol{\beta})=\boldsymbol{w}_n^\tau M_n\boldsymbol{w}_n$$

$$M_n=\sigma_a^2\Gamma_n^{-1}=\sigma_a^2(E\boldsymbol{w}_n\boldsymbol{w}_n^\tau)^{-1}$$

由于 w_t 可以用传递形式解出，即

$$w_t=\varphi^{-1}(B)\theta(B)a_t=\psi(B)a_t=\sum_{j=0}^{\infty}\psi_j a_{t-j}$$

其中 $\psi_0 = 1$，于是

$$Ea_k w_t = \sum_{j=0}^{\infty} \psi_j E a_k a_{t-j} = \begin{cases} \sigma_a^2 \psi_{t-k}, & (t \geqslant k) \\ 0, & (t < k) \end{cases}$$

若形式地记 $\psi_l = 0(l < 0)$，则无论 t 是否小于 k，可统一表

$$Ea_k w_t = \sigma_a^2 \psi_{t-k}$$

而且当 $1 \leqslant i, j \leqslant n$ 时，有

$$\gamma_{i-j} = E w_i w_j = \sigma_a^2 \sum_{k=0}^{\infty} \psi_k \psi_{k+j-i} = \sigma_a^2 \sum_{l=-\infty}^{n} \psi_{i-l} \psi_{j-l} \quad (3.2)$$

由此便得到

$$\sum_{k=-\infty}^{n} E(a_k \boldsymbol{w}_n) E(a_k \boldsymbol{w}_n^{\tau}) = \sigma_a^2 \Gamma_n$$

从而有

$$\begin{aligned} \sum_{k=-\infty}^{n} \hat{a}_k^2 &= \sum_{k=-\infty}^{n} [E(a_k | \boldsymbol{w}_n)]^2 \\ &= \sum_{k=-\infty}^{n} \boldsymbol{w}_n^{\tau} \Gamma_n^{-1} E(a_k \boldsymbol{w}_n) E(a_k \boldsymbol{w}_n^{\tau}) \Gamma_n^{-1} \boldsymbol{w}_n \\ &= \boldsymbol{w}_n^{\tau} \Gamma_n^{-1} \sum_{k=-\infty}^{n} E(a_k \boldsymbol{w}_n) E(a_k \boldsymbol{w}_n^{\tau}) \Gamma_n^{-1} \boldsymbol{w}_n \\ &= \sigma_a^2 \boldsymbol{w}_n^{\tau} \Gamma_n^{-1} \Gamma_n \Gamma_n^{-1} \boldsymbol{w}_n = \sigma_a^2 \boldsymbol{w}_n^{\tau} \Gamma_n^{-1} \boldsymbol{w}_n \\ &= \boldsymbol{w}_n^{\tau} M_n \boldsymbol{w}_n = S_n(\boldsymbol{\beta}) \end{aligned} \quad (3.3)$$

因此，我们在本书中常称 $S_n(\boldsymbol{\beta})$ 为平方和函数.

如果 w_t 为 MA(o, p) 序列，则

$$S_n(\boldsymbol{\beta}) = S_n(\boldsymbol{\theta}) = \sum_{k=-q+1}^{n} \hat{a}_k^2 \quad (3.4)$$

如果 w_t 为 AR$(p, 0)$ 序列，我们还可以进一步求出 $S_n(\boldsymbol{\beta}) = S_n(\boldsymbol{\varphi})$ 的解析形式. 为此，对于 $n > p$，令

$$T_n = \begin{pmatrix} & & 1 \\ 0 & 1 & 0 \\ & \cdot^{\cdot^{\cdot}} & \\ 1 & & \end{pmatrix}_{n\times n}, \qquad P_n = \begin{pmatrix} I_p & & & 0 \\ -\varphi_p \cdots -\varphi_1 & & 1 & \\ 0 \quad \ddots & & \ddots & \ddots \\ & -\varphi_p \cdots -\varphi_1 & & 1 \end{pmatrix}_{n\times n}$$

(3.5)

于是依次可得以下诸式:

$$P_n \boldsymbol{w}_n = (w_1, w_2, \cdots, w_p, a_{p+1}, \cdots, a_n)^\tau$$

$$P_n(E\boldsymbol{w}_n\boldsymbol{w}_n^\tau)P_n^\tau = E(P_n\boldsymbol{w}_n\boldsymbol{w}_n^\tau P_n^\tau) = \begin{pmatrix} E\boldsymbol{w}_p\boldsymbol{w}_p^\tau & 0 \\ 0 & \sigma_a^2 I_{(n-p)} \end{pmatrix}$$

$$= \sigma_a^2 \begin{pmatrix} M_p^{-1} & 0 \\ 0 & I_{(n-p)} \end{pmatrix}$$

$$M_n = \sigma_a^2(E\boldsymbol{w}_n\boldsymbol{w}_n^\tau)^{-1} = P_n^\tau \begin{pmatrix} M_p & 0 \\ 0 & I \end{pmatrix} P_n \tag{3.6}$$

$$S_n(\boldsymbol{\varphi}) = \boldsymbol{w}_n^\tau M_n \boldsymbol{w}_n = \boldsymbol{w}_n^\tau P_n^\tau \begin{pmatrix} M_p & 0 \\ 0 & I \end{pmatrix} P_n \boldsymbol{w}_n$$

$$= \boldsymbol{w}_p^\tau M_p \boldsymbol{w}_p + \sum_{k=p+1}^{n} (w_k - \varphi_1 w_{k-1} - \cdots - \varphi_p w_{k-p})^2 \tag{3.7}$$

$$T_n = T_n^{-1} = T_n^\tau$$

再令

$$(w_1^*, w_2^*, \cdots, w_n^*)^\tau = \boldsymbol{w}_n^* = T_n \boldsymbol{w}_n$$

$$= (w_n, w_{n-1}, \cdots, w_1)^\tau$$

即 $w_k^* = w_{n-k+1}(1 \leqslant k \leqslant n)$. 由 w_t 的平稳性易知,

$$T_n(E\boldsymbol{w}_n\boldsymbol{w}_n^\tau)T_n^\tau = E\boldsymbol{w}_n^*(\boldsymbol{w}_n^*)^\tau = E\boldsymbol{w}_n\boldsymbol{w}_n^\tau,$$

于是

$$T_n^\tau M_n T_n = M_n$$

$$S_n(\boldsymbol{\varphi}) = \boldsymbol{w}_n^\tau M_n \boldsymbol{w} = \boldsymbol{w}_n^\tau T_n^\tau M_n T_n \boldsymbol{w}_n = \boldsymbol{w}_n^* M_n w_n^*$$

$$= \boldsymbol{w}_p^{*\tau} M_p \boldsymbol{w}_p^* + \sum_{k=p+1}^{n} (w_k^* - \varphi_1 w_{k-1}^* - \cdots - \varphi_p w_{k-p}^*)^2 \tag{3.8}$$

其中 $\boldsymbol{w}_p^* = (w_n, w_{n-1}, \cdots, w_{n-p+1})^\tau$. 由此可知,对于任意 n 维向量 $\boldsymbol{w}_n$, 成立恒等式

$$\boldsymbol{w}_p^{*\tau} M_p \boldsymbol{w}_p^* + \sum_{k=p+1}^{n} (w_{n-k+1} - \varphi_1 w_{n-k+2} - \cdots - \varphi_p w_{n-k+p+1})^2$$

$$= \boldsymbol{w}_p^\tau M_p \boldsymbol{w}_p + \sum_{k=p+1}^{n} (w_k - \varphi_1 w_{k-1} - \cdots - \varphi_p w_{k-p})^2 \tag{3.9}$$

我们先考虑一种特殊情况，即取 $n=2p$，$w_{2p}^{\tau}=(w_1, w_2, \cdots, w_p, 0, 0, \cdots, 0)$，这时，$w_{2p}^{*\tau}=(0, \cdots, 0, w_p, w_{p-1}, \cdots, w_1)$. 以此代入(3.9)式并移项可得

$$w_p^{\tau} M_p w_p = \sum_{k=p+1}^{2p}(w_{2p-k+1}-\varphi_1 w_{2p-k+2}-\cdots-\varphi_p w_{2p-k+p+1})^2 - \sum_{k=p+1}^{2p}(w_k-\varphi_1 w_{k-1}-\cdots-\varphi_p w_{k-p})^2$$

注意此式中 $w_p^{\tau}=(w_1, w_2, \cdots, w_p)$，$w_k=0\ (k>p)$. 暂记 $\varphi_0=-1$，上式又可按如下方式改写：

$$\begin{aligned}
& w_p^{\tau} M_p w_p \\
&= \sum_{l=1}^{p}(w_{p-l+1}-\varphi_1 w_{p-l+2}-\cdots-\varphi_p w_{p-l+p+1})^2 \\
&\quad - \sum_{l=1}^{p}(w_{l+p}-\varphi_1 w_{l+p-1}-\cdots-\varphi_p w_l)^2 \\
&= \sum_{l=1}^{p}\left[\left(\sum_{i=0}^{p}\varphi_i w_{p-l+1+i}\right)^2-\left(\sum_{i=0}^{p}\varphi_i w_{l+p-i}\right)^2\right] \\
&= \sum_{l=1}^{p}\sum_{i,j=0}^{p}\varphi_i\varphi_j(w_{p-l+1+i}w_{p-l+1+j}-w_{l+p-i}w_{l+p-j}) \\
&= \sum_{i,j=0}^{p}\varphi_i\varphi_j\sum_{k=1}^{p}(w_{p-l+1+i}w_{p-l+1+j}-w_{l+p-i}w_{l+p-j}) \\
&= \sum_{i,j=0}^{p}\varphi_i\varphi_j\left(\sum_{s=1+j}^{p+j}w_{s+i-j}w_s-\sum_{s=p+1-i}^{2p-i}w_s w_{s+i-j}\right) \\
&= \left\{\sum_{\substack{i,j=0\\ i+j=p}}^{p}+\sum_{\substack{i,j=0\\ i+j<p}}^{p}+\sum_{\substack{i,j=0\\ i+j>p}}^{p}\right\}\varphi_i\varphi_j\left(\sum_{s=1+j}^{p}w_{s+i-j}w_s\right. \\
&\quad \left.-\sum_{s=p+1-i}^{p}w_{s+i-j}w_s\right) \quad (\text{注意 } w_s=0,\ s>p) \\
&= \sum_{\substack{i,j=0\\ i+j<p}}^{p}\varphi_i\varphi_j\sum_{s=1+j}^{p-i}w_{s+i-j}w_s-\sum_{\substack{i,j=0\\ i+j>p}}^{p}\varphi_i\varphi_j\sum_{s=p+1-i}^{j}w_{s+i-j}w_s
\end{aligned} \tag{3.10}$$

由此可以求出当 $n>p$ 时 $S_n(\varphi)$ 的一般表达式．事实上，由 (3.7) 和 (3.10) 式

$$
\begin{aligned}
S_n(\varphi) &= w_p^\tau M_p w_p + \sum_{k=p+1}^{n} (w_k - \varphi_1 w_{k-1} - \cdots - \varphi_p w_{k-p})^2 \\
&= w_p^\tau M_p w_p + \sum_{i,j=0}^{p} \varphi_i \varphi_j \sum_{k=p+1}^{n} w_{k-i} w_{k-j} \\
&- \sum_{\substack{i,j=0 \\ i+j<p}}^{p} \varphi_i \varphi_j \sum_{s=1+j}^{p-i} w_{s+i-j} w_s \\
&\quad - \sum_{\substack{i,j=0 \\ i+j>p}}^{p} \varphi_i \varphi_j \sum_{s=p+1-i}^{j} w_{s+i-j} w_s \\
&\quad + \sum_{i,j=0}^{p} \varphi_i \varphi_j \sum_{s=p+1-i}^{n-i} w_s w_{s+i-j}
\end{aligned}
$$

注意将上式右边最后一求和项按照 $i+j\leqslant p$，$p<i+j\leqslant n$ 和 $i+j>n$ 三种情况分别求和，即可得到

$$
\begin{aligned}
S_n(\varphi) = & \sum_{\substack{i,j=0 \\ i+j\leqslant p}}^{p} \varphi_i \varphi_j \sum_{s=1+j}^{n-i} w_{s+i-j} w_s \\
& + \sum_{\substack{i,j=0 \\ p<i+j\leqslant n}}^{p} \varphi_i \varphi_j \sum_{s=1+j}^{n-i} w_{s+i-j} w_s \\
& - \sum_{\substack{i,j=0 \\ n<i+j}}^{p} \varphi_i \varphi_j \sum_{s=n-i+1}^{j} w_{s+i-j} w_s \\
= & \sum_{\substack{i,j=0 \\ i+j\leqslant n}}^{p} \varphi_i \varphi_j \sum_{s=1+j}^{n-i} w_{s+i-j} w_s \\
& - \sum_{\substack{i,j=0 \\ i+j>n}}^{p} \varphi_i \varphi_j \sum_{s=n-i+1}^{j} w_{s+i-j} w_s
\end{aligned} \tag{3.11}
$$

参照 (3.10) 式，可知上式对 $n\geqslant p$ 都成立．特别，当 $n\geqslant 2p$ 时，由于 $0\leqslant i,\ j\leqslant p$，故 $i+j\leqslant 2p\leqslant n$，所以上式右边第二求和项 $\sum\limits_{i+j>n}^{p}$ 不复存在．因此这时有

$$
\begin{aligned}
S_n(\boldsymbol{\varphi}) = & \sum_{i,j=0}^{p} \varphi_i \varphi_j \sum_{s=1+j}^{n-i} w_{s+i-j} w_s \\
& - \sum_{i,j=0}^{p} \varphi_i \varphi_j \sum_{t=1+i+j}^{n} w_{t-i} w_{t-j}
\end{aligned} \tag{3.12}
$$

若记

$$
d_{ij} = \sum_{s=1+j}^{n-i} w_{s+i-j} w_s \qquad (0 \leqslant i,\ j \leqslant p)
$$

则 (3.12) 式又可写做 (注意 $\varphi_0 = -1$)

$$
\begin{aligned}
S_n(\boldsymbol{\varphi}) = & \sum_{i,j=0}^{p} \varphi_i \varphi_j d_{ij} = d_{00} - 2 \sum_{j=1}^{p} \varphi_j d_{0j} \\
& + \sum_{i,j=1}^{p} \varphi_i \varphi_j d_{ij}, \quad n \geqslant 2p
\end{aligned} \tag{3.13}
$$

(3.10), (3.11) 和 (3.12) 式都是参数 $\varphi_i (0 \leqslant i \leqslant p)$ 的二次型函数. 此外,还可以给出 $S_n(\boldsymbol{\varphi})$ 的另一种有用的表达式. 为此,我们仍以 $n = 2p$, $\boldsymbol{w}_{2p}^{\tau} = (\boldsymbol{w}_p^{\tau}, 0, \cdots 0)$ 代入 (3.9) 式, 并且暂记 $\varphi_{-i} = \varphi_{p+i} = 0$ $(i > 0)$, 于是可得

$$
\begin{aligned}
\boldsymbol{w}_p^{\tau} M_p \boldsymbol{w}_p = & \sum_{l=1}^{p} [(w_{p-l+1} - \varphi_1 w_{p-l+2} - \cdots - \varphi_p w_{p-l+p+1})^2 \\
& - (w_{p+l} - \varphi_1 w_{p+l-1} - \cdots - \varphi_p w_l)^2] \\
& \qquad\qquad (\text{注意 } w_s = 0,\ s > p) \\
= & \sum_{l=1}^{p} \left[\Big(\sum_{s=p-l+1}^{p} \varphi_{s-p+l-1} w_s \Big)^2 - \Big(\sum_{s=l}^{p} \varphi_{p-s+l} w_s \Big)^2 \right] \\
= & \sum_{l=1}^{p} \left[\Big(\sum_{s=1}^{p} \varphi_{s-p+l-1} w_s \Big)^2 - \Big(\sum_{s=1}^{p} \varphi_{p-s+l} w_s \Big)^2 \right] \\
& \qquad\qquad (\text{注意 } \varphi_{-i} = \varphi_{p+i} = 0,\ i > 0) \\
= & \sum_{t,s=1}^{p} w_s w_t \sum_{l=1}^{p} (\varphi_{s-p+l-1} \varphi_{t-p+l-1} - \varphi_{p-s+l} \varphi_{p-t+l}) \\
= & \sum_{t,s=1}^{p} w_s w_t \Big(\sum_{i=\max(0,s-t)}^{s-1} \varphi_i \varphi_{i+t-s}
\end{aligned}
$$

$$- \sum_{i=p-t+1}^{\min(p,p+s-t)} \varphi_i \varphi_{i+t-s} \Big)$$

$$= \sum_{s,t=1}^{p} w_s w_t \Big(\sum_{i=\max(0,s-t)}^{\min(s-1,p-t)} \varphi_i \varphi_{i+t-s}$$

$$- \sum_{i=\max(s,p-t+1)}^{\min(p,p+s-t)} \varphi_i \varphi_{i+t-s} \Big) \tag{3.14}$$

若以 $m_{st}^{(p)}$ 代表矩阵 M_p 的 s 行 t 列元，由上式有

$$m_{st}^{(p)} = \sum_{i=\max(0,s-t)}^{\min(s-1,p-t)} \varphi_i \varphi_{i+t-s} - \sum_{i=\max(s,p-t+1)}^{\min(p,p+s-t)} \varphi_i \varphi_{i+t-s} \tag{3.15}$$

对于 $n > p$，由 (3.7) 式可知

$$w_n^\tau M_n w_n = S_n(\varphi) = w_p^\tau M_p w_p$$

$$+ \sum_{k=p+1}^{n} (w_k - \varphi_1 w_{k-1} - \cdots - \varphi_p w_{k-p})^2$$

$$= w_p^\tau M_p w_p + \sum_{k=p+1}^{n} \Big(\sum_{s=k-p}^{k} \varphi_{k-s} w_s \Big)^2$$

$$= w_p^\tau M_p w_p + \sum_{k=p+1}^{n} \sum_{s,t=k-p}^{k} \varphi_{k-s} \varphi_{k-t} w_s w_t$$

$$= w_p^\tau M_p w_p + \sum_{k=p+1}^{n} \sum_{s,t=1}^{n} \varphi_{k-s} \varphi_{k-t} w_s w_t$$

$$(\text{记 } \varphi_{-i} = \varphi_{p+i} = 0 \quad i > 0)$$

$$= w_p^\tau M_p w_p + \sum_{s,t=1}^{n} w_s w_t \sum_{k=p+1}^{n} \varphi_{k-s} \varphi_{k-t}$$

$$= w_p M_p w_p + \sum_{s,t=1}^{n} w_s w_t \sum_{i=\max(p+1-t,0,s-t)}^{\min(n-t,p,p+s-t)} \varphi_{i+t-s} \varphi_i \tag{3.16}$$

由此式可以得到 $M_n(n > p)$ 的诸元 $m_{st}^{(n)}$ 的表达式：

$$m_{st}^{(n)} = \begin{cases} m_{st}^{(p)} + \displaystyle\sum_{i=p+1-t}^{\min(n-t,p,p+s-t)} \varphi_{i+t-s} \varphi_i & 1 \leqslant s,\ t \leqslant p \\ \displaystyle\sum_{i=\max(p+1-t,0,t-t)}^{\min(n-t,p,p+s-t)} \varphi_{i+t-s} \varphi_i, & p < \max(s,t) \leqslant n \end{cases} \tag{3.17}$$

从而也就获得了 AR $(p, 0)$ 序列 w_t 的自协方差阵 Γ_n 之逆 $\Gamma_n^{-1} = \sigma_a^{-2} M_n$ 的表达式. 这是本节的一个附带收获. (3.17) 式还告诉我们,当 $|s-t| > p$ 时, Γ_n^{-1} (或 M_n) 的 s 行 t 列元全为零, 而其余元都是 $\varphi_i (0 \leqslant i \leqslant p)$ 的二次型函数.

由 P_n 的定义 (见 (3.5) 式) 易知其行列式 $|P_n| = 1$, 于是由 (3.6) 式可知

$$|M_n| = |P_n^{\tau}| |M_p| |P_n| = |M_p|, \quad n \geqslant p$$

到此为止, 我们已将对应于 AR $(p, 0)$ 序列中 M_n 的性质分析清楚,而且给出了 $S_n(\boldsymbol{\varphi})$ 的具体表达式. 这样, $\boldsymbol{w}_n$ 的分布密度函数和相应的对数似然函数也就有了清楚的表示:

$$p(\boldsymbol{w}_n | \boldsymbol{\varphi}, \sigma_a^2) = (2\pi\sigma_a^2)^{-\frac{n}{2}} |M_p|^{\frac{1}{2}} \exp\left\{-\frac{S_n(\boldsymbol{\varphi})}{2\sigma_a^2}\right\} \tag{3.18}$$

$$l(\boldsymbol{\varphi}, \sigma_a^2 | \boldsymbol{w}_n) = -\frac{n}{2}\log 2\pi - \frac{1}{2}\left\{n \log \sigma_a^2 - \log |M_p| + \frac{S_n(\boldsymbol{\varphi})}{\sigma_a^2}\right\} \tag{3.19}$$

其中 M_p 和 $S_n(\boldsymbol{\varphi})$ 的表达式分别见 (3.15) 式和 (3.11)(或(3.12)) 式.

下面我们来讨论一般的 ARMA 序列的分布密度函数. 假定 w_t 是正态 ARMA (p, q) 序列,

$$\varphi(B) w_t = \theta(B) a_t$$

由此可以将 $a_k (1 \leqslant k \leqslant n)$ 依次表成 $a_j(-q+1 \leqslant j \leqslant 0)$ 和 $w_k(-p+1 \leqslant k \leqslant n)$ 的线性关系式:

$$\begin{aligned}
a_1 &= w_1 - \varphi_1 w_0 - \cdots - \varphi_p w_{1-p} + \theta_1 a_0 + \cdots + \theta_q a_{1-q} \\
a_2 &= w_2 - \varphi_1 w_1 - \cdots - \varphi_p w_{2-p} + \theta_1 a_1 + \cdots + \theta_q a_{2-q} \\
&= w_2 - (\varphi_1 - \theta_1) w_1 - (\varphi_2 + \theta_1 \varphi_1) w_0 - \cdots \\
&\quad - (\varphi_p + \theta_1 \varphi_{p-1}) w_{2-p} - \theta_1 \varphi_p w_{1-p} + (\theta_1^2 + \theta_2) a_0 \\
&\quad + (\theta_1 \theta_2 + \theta_3) a_{-1} + \cdots + (\theta_1 \theta_{q-1} \\
&\quad + \theta_q) a_{2-q} + \theta_1 \theta_q a_{1-q}
\end{aligned}$$

$\cdots\cdots\cdots$

记

$$\boldsymbol{a}_* = (a_{-q+1}, a_{-q+2}, \cdots, a_0)^\tau$$
$$\boldsymbol{w}_* = (w_{-p+1}, w_{-p+2}, \cdots, w_0)^\tau$$
$$\boldsymbol{\xi}^\tau = (\boldsymbol{a}_*^\tau, \boldsymbol{w}_*^\tau)$$
$$\boldsymbol{a}_n = (a_1, a_2, \cdots, a_n)^\tau$$
$$\boldsymbol{w}_n = (w_1, w_2, \cdots, w_n)^\tau$$

则上面的一组关系式又可写做

$$\boldsymbol{a}_n = X_n\boldsymbol{\xi} + L_n\boldsymbol{w}_n \tag{3.20}$$

其中 X_n 和 L_n 分别为 $n \times (p+q)$ 和 $n \times n$ 阶矩阵，它们是由上面诸式中的系数决定的，其元为 $\boldsymbol{\varphi}$ 和 $\boldsymbol{\theta}$ 的函数，特别 L_n 是对角线元全等于 1 的下三角阵(因而是可逆阵)，此外，我们并不详细推导它们的具体形式．由(3.20)式可知，

$$\boldsymbol{w}_n = L_n^{-1}(\boldsymbol{a}_n - X_n\boldsymbol{\xi}) \tag{3.21}$$

对于式中诸随机量，在已知 $\boldsymbol{w}_n$ 的条件下分别求它们的最小方差估计，可得

$$\boldsymbol{w}_n = L_n^{-1}(\hat{\boldsymbol{a}}_n - X_n\hat{\boldsymbol{\xi}}) \tag{3.22}$$

其中

$$\hat{\boldsymbol{a}}_n = E(\boldsymbol{a}_n|\boldsymbol{w}_n), \qquad \hat{\boldsymbol{\xi}} = E(\boldsymbol{\xi}|\boldsymbol{w}_n)$$

若又记

$Q = \dfrac{1}{\sigma_a^2}E\boldsymbol{\xi}\boldsymbol{\xi}^\tau$，应用(3.21)式和 $\boldsymbol{\xi}$ 与 $\boldsymbol{a}_n$ 的相互独立性，并注意使用矩阵求逆公式(见[20])，有

$$\begin{aligned}\sigma_a^{-2}M_n &= [E\boldsymbol{w}_n\boldsymbol{w}_n^\tau]^{-1} = [EL_n^{-1}(\boldsymbol{a}_n - X_n\boldsymbol{\xi}) \\ &\quad \cdot (\boldsymbol{a}_n^\tau - \boldsymbol{\xi}^\tau X_n^\tau)L_n^{-\tau}]^{-1} \\ &= \sigma_a^{-2}[L_n^\tau(I + X_nQX_n^\tau)^{-1}L_n] \\ &= \sigma_a^{-2}L_n^\tau[I - X_n(Q^{-1} + X_n^\tau X_n)^{-1}X_n^\tau]L_n \end{aligned} \tag{3.23}$$

由此并利用(3.21)式和最小方差估计的性质，可得

$$\begin{aligned}\hat{\boldsymbol{\xi}} &= (E\boldsymbol{\xi}\boldsymbol{w}_n^\tau)(E\boldsymbol{w}_n\boldsymbol{w}_n^\tau)^{-1}\boldsymbol{w}_n \\ &= [E\boldsymbol{\xi}(\boldsymbol{a}_n^\tau - \boldsymbol{\xi}^\tau X_n^\tau)L_n^{-\tau}][E\boldsymbol{w}_n\boldsymbol{w}_n^\tau]^{-1}\boldsymbol{w}_n \\ &= -QX_n^\tau L_n^{-\tau}L_n^\tau[I - X_n(Q^{-1} + X_n^\tau X_n)^{-1}X_n^\tau]L_n\boldsymbol{w}_n \\ &= -Q[I - X_n^\tau X_n(Q^{-1} + X_n^\tau X_n)^{-1}]X_n^\tau L_n\boldsymbol{w}_n\end{aligned}$$

$$= -Q[Q^{-1} + X_n^\tau X_n - X_n^\tau X_n](Q^{-1} + X_n^\tau X_n)^{-1} X_n^\tau L_n \boldsymbol{w}_n$$
$$= -(Q^{-1} + X_n^\tau X_n)^{-1} X_n^\tau L_n \boldsymbol{w}_n \tag{3.24}$$

类似地有

$$\hat{\boldsymbol{a}}_n = [I - X_n(Q^{-1} + X_n^\tau X_n)^{-1} X_n^\tau] L_n \boldsymbol{w}_n$$

将(3.22)式代入(3.24)式，并以 $(Q^{-1} + X_n^\tau X_n)$ 左乘两边，又有

$$(Q^{-1} + X_n^\tau X_n)\hat{\boldsymbol{\xi}} = -X_n^\tau L_n L_n^{-1}(\hat{\boldsymbol{a}}_n - X_n \hat{\boldsymbol{\xi}})$$
$$= -X_n^\tau \hat{\boldsymbol{a}}_n + X_n^\tau X_n \hat{\boldsymbol{\xi}}$$

于是

$$Q^{-1}\hat{\boldsymbol{\xi}} = -X_n^\tau \hat{\boldsymbol{a}}_n \quad 即 \quad \hat{\boldsymbol{a}}_n^\tau X_n = -\hat{\boldsymbol{\xi}}^\tau Q^{-1} \tag{3.25}$$

最后，联合使用(3.22)，(3.23)和(3.25)式，即得

$$S_n(\boldsymbol{\beta}) = \boldsymbol{w}_n^\tau M_n \boldsymbol{w}_n$$
$$= (\hat{\boldsymbol{a}}_n^\tau - \hat{\boldsymbol{\xi}}^\tau X_n^\tau)[I - X_n(Q^{-1} + X_n^\tau X_n)^{-1} X_n^\tau] L_n \boldsymbol{w}_n$$
$$= (\hat{\boldsymbol{a}}_n^\tau - \hat{\boldsymbol{\xi}}^\tau X_n^\tau)\hat{\boldsymbol{a}}_n = \hat{\boldsymbol{a}}_n^\tau \hat{\boldsymbol{a}}_n + \hat{\boldsymbol{\xi}}^\tau Q^{-1} \hat{\boldsymbol{\xi}}$$
$$= (\hat{\boldsymbol{\xi}}^\tau,\ \hat{\boldsymbol{a}}_n^\tau)\begin{pmatrix} Q^{-1} & 0 \\ 0 & I_n \end{pmatrix}\begin{pmatrix} \hat{\boldsymbol{\xi}} \\ \hat{\boldsymbol{a}}_n \end{pmatrix} \tag{3.26}$$

注意使用求分块矩阵的行列式的公式，即如果分块方阵

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

中，子阵 A_{11} 和 A_{22} 可逆，则行列式

$$|A| = |A_{22}||A_{11} - A_{12}A_{22}^{-1}A_{21}| = |A_{11}||A_{22} - A_{21}A_{11}^{-1}A_{12}| \tag{3.27}$$

取 $A_{22} = I_n$，$A_{11} = Q^{-1}$，$A_{21} = X$，$A_{12} = -X^\tau$，可知

$$|Q^{-1}||I_n + X_n Q X_n^\tau| = |Q^{-1} + X_n^\tau X_n|$$

又由(3.23)式的中间关系式可知

$$|M_n| = |L_n||I + X_n Q X_n^\tau|^{-1}|L_n| = |I + X_n Q X_n^\tau|^{-1}$$

其中由于 L_n 是对角线元全等于1的下三角阵，所以 $|L_n| = 1$。再利用前一式即得

$$|M_n| = \frac{|Q^{-1}|}{|Q^{-1} + X_n^\tau X_n|} \tag{3.28}$$

这样一来，ARMA (p, q) 正态序列的分布密度函数和对数似然

函数分别可以写成

$$p(\boldsymbol{w}_n|\boldsymbol{\beta}, \sigma_a^2) = (2\pi\sigma_a^2)^{-\frac{n}{2}} \frac{|Q^{-1}|^{\frac{1}{2}}}{|Q^{-1} + X_n^\tau X_n|^{\frac{1}{2}}} \cdot \exp\left\{-\frac{1}{2\sigma_a^2}(\boldsymbol{\xi}^\tau Q^{-1}\boldsymbol{\xi} + \hat{\boldsymbol{a}}_n^\tau \hat{\boldsymbol{a}}_n)\right\} \quad (3.29)$$

$$l(\boldsymbol{\beta}, \sigma_a^2|\boldsymbol{w}_n) = -\frac{n}{2}\log 2\pi - \frac{1}{2}\left\{n\log\sigma_a^2 - \log|Q^{-1}| + \log|Q^{-1} + X_n^\tau X_n| + \frac{1}{\sigma_a^2}(\boldsymbol{\xi}^\tau Q^{-1}\boldsymbol{\xi} + \hat{\boldsymbol{a}}_n^\tau \hat{\boldsymbol{a}}_n)\right\} \quad (3.30)$$

根据(3.3)式又知

$$\boldsymbol{\xi}^\tau Q^{-1}\boldsymbol{\xi} = \sum_{k=-\infty}^{0} \hat{a}_k^2 = \sum_{k=-\infty}^{0} [E(a_k|\boldsymbol{w}_n)]^2 \quad (3.31)$$

由此可见上式左边的级数是均方收敛的.

现在补充说明 $Q = \frac{1}{\sigma_a^2} E\boldsymbol{\xi}\boldsymbol{\xi}^\tau$ 是可逆阵.如果不然,必有 $|Q| = 0$, 即存在非全为零的实数组 $(\lambda_0, \lambda_1, \cdots, \lambda_{p-1}, \mu_0, \mu_1, \cdots, \mu_{q-1})$ 使的

$$\lambda_0 w_0 + \lambda_1 w_{-1} + \cdots + \lambda_{p-1} w_{-p+1} = \mu_0 a_0 + \mu_1 a_{-1} + \cdots + \mu_{q-1} a_{-q+1}$$

由于 w_t 和 a_t 是平稳正态序列,所以,事实上对任意的 t, 都有

$$\lambda_0 w_t + \lambda_1 w_{t-1} + \cdots + \lambda_{p-1} w_{t-p+1} = \mu_0 a_t + \mu_1 a_{t-1} + \cdots + \mu_{q-1} a_{t-q+1}$$

或者令 $\Lambda(B) = \sum_{j=0}^{p-1} \lambda_j B^j$, $\mu(B) = \sum_{j=0}^{q-1} \mu_j B^j$, 上式又可写做

$$\Lambda(B) w_t = \mu(B) a_t$$

由于 a_t 是白噪声序列,所以 $(\lambda_0, \lambda_1, \cdots, \lambda_{p-1})$ 不能全部为零,否则必然 $\mu_0, \mu_1, \cdots, \mu_{q-1}$ 也全部为零,而这与 $\lambda_i, \mu_j (0 \leqslant i \leqslant p-1, 0 \leqslant j \leqslant q-1)$ 不全为零的假定不符. 由此可见, $\Lambda(B)$ 为非零的多项式. 又由于 w_t 为 ARMA 序列, 满足 $\varphi(B)w_t = \theta(B)a_t$, 所以

$$\Lambda(B)\theta(B)a_t = \varphi(B)\Lambda(B)w_t = \mu(B)\varphi(B)a_t$$

由此可知 $A(B)\theta(B)=\mu(B)\varphi(B)$. 由于 $\varphi(B)$ 有 p 个根，$\theta(B)$ 有 q 个根，而两组根中又没有公共的根，因此这个恒等式表明 $A(B)$ 与 $\varphi(B)$ 的根必须完全一致，$\theta(B)$ 与 $\mu(B)$ 的根也必须完全一致. 这是不可能的，因为非零多项式 $A(B)$ 最多只有 $p-1$ 个根（包括重根）. 这就用反证法证明了 Q 的可逆性.

从前面已经知道，对于 AR $(p, 0)$ 序列，当 $n \geqslant p$ 时，$|M_n|=|M_p|$，因此，它与 n 无关；但是，对于一般的 ARMA 序列，$|M_n|=|Q^{-1}||Q^{-1}+X_n^{\tau}X_n|^{-1}$ 通常与 n 有关，而且这时我们给不出 M_n 的无的一般表达式. 下面我们来证明 $|M_n|$ 有非零的上下界，这性质以后将要用到.

引理. 若 A, B 都是对称正定阵，而且 $A \leqslant B$，那么 $|A| \leqslant |B|$.

证明. 因为 A, B 都是正定阵，所以必存在可逆阵 P，使 $P^{\tau}AP$ 为单位阵，$P^{\tau}BP$ 为对角阵. 由 $A \leqslant B$ 知 $P^{\tau}AP \leqslant P^{\tau}BP$，显然这时有 $1=|P^{\tau}AP| \leqslant |P^{\tau}BP|$. 于是 $|A|=|P|^{-2} \leqslant |P|^{-2}|P^{\tau}BP| = |P|^{-2}|P|^{2}|B|=|B|$.

定理 1. 若 w_t 为正态 ARMA 序列，那么存在一个与参数有关而与 n 无关的正数 C_1，使得对一切 n，

$$0<C_1 \leqslant |M_n|=\frac{|Q^{-1}|}{|Q^{-1}+X_n^{\tau}X_n|} \leqslant 1 \tag{3.32}$$

证明. 在 (3.20) 式中，特别取 $w_k=0\ (k \geqslant 1)$，则相应的 $\boldsymbol{a}_n$ 为

$$\boldsymbol{a}_n=X_n\boldsymbol{\xi}+L_n\boldsymbol{w}_n=X_n\boldsymbol{\xi} \tag{3.33}$$

而 $\boldsymbol{a}_n$ 的分量 a_k 满足递推式

$a_k=w_k-\varphi_1 w_{k-1}-\varphi_2 w_{k-2}-\cdots-\varphi_p w_{k-p}+\theta_1 a_{k-1}+\cdots+\theta_q a_{k-q}$，$k \geqslant 1$ 于是当 $k \geqslant p+1$ 时，由于取 $\boldsymbol{w}_n=0$，所以有

$$a_k=\theta_1 a_{k-1}+\theta_2 a_{k-2}+\cdots+\theta_q a_{k-q}, \quad k \geqslant p+1 \tag{3.34}$$

即 $a_k(k \geqslant p+1)$ 是齐次差分方程 $\theta(B)a_k=0$ 的解. 由于 $\theta(B)$ 的根都在单位圆外，由差分方程的稳定性理论可知，a_k 与初

始值 $(a_p, a_{p-1}, \cdots, a_{p-q+1}) = \eta^\tau$ 之间满足以下不等式:

$$a_k^2 \leqslant K_1\mu^{2k}(\eta^\tau\eta)$$

其中 K_1 和 μ 是依 $\theta(B)$ 的根而确定的与 η 和 k 无关的正常数，而且 $0<\mu<1$. 再由 (3.20) 式的定义可知，η 和 ξ 有线性关系，所以

$$\eta^\tau\eta \leqslant K_2\xi^\tau\xi$$

于是，由 (3.33) 式，对任意的 ξ 和 n,

$$\xi^\tau X_n^\tau X_n\xi = \alpha_n^\tau\alpha_n = \sum_{k=1}^{n} a_k^2 \leqslant K_1K_2 \sum_{k=1}^{n} \mu^{2k}(\xi^\tau\xi)$$

$$\leqslant K_1K_2(\xi^\tau\xi) \sum_{k=0}^{\infty} \mu^{2k} = K_3(\xi^\tau\xi)$$

即存在正常数 K_3，使得对任何 n,

$$X_n^\tau X_n \leqslant K_3 I_{p+q}$$

再由引理即知

$$|Q^{-1}| \leqslant |Q^{-1} + X_n^\tau X_n| \leqslant |Q^{-1} + K_3 I_{p+q}| = K_4 < \infty$$

于是便证明了，对任何 n,

$$1 \geqslant |M_n| = \frac{|Q^{-1}|}{|Q^{-1} + X_n^\tau X_n|} \geqslant \frac{|Q^{-1}|}{K_4} = C_1 > 0$$

此即 (3.32) 式. 由此可得

$$0 \geqslant \log|M_n| \geqslant \log C_1 = -C_2 > -\infty \tag{3.35}$$

其中 $C_2 = -\log C_1 > 0$. 此式说明 $|\log|M_n||$ 亦有界.

下面我们还将证明 $\left|\frac{\partial \log|M_n|}{\partial\beta_i}\right|$ 对 n 而言也有界，这一性质也是今后要用到的.

定理 2. 在与定理 1 相同的条件下，必有常数 $C_3(>0)$，使得对一切 n,

$$\left|\frac{\partial \log|M_n|}{\partial\beta_i}\right| \leqslant C_3 \tag{3.36}$$

证明. 如同在定理 1 的证明中那样，在 (3.20) 式中取 $w_n = 0$，于是成立 (3.33) 和 (3.34) 式. 现在对这些公式求偏导数:

$$\frac{\partial \alpha_n}{\partial \beta_i} = \frac{\partial X_n}{\partial \beta_i} \xi \tag{3.37}$$

$$\frac{\partial a_k}{\partial \beta_i} = \begin{cases} \theta_1 \frac{\partial a_{k-1}}{\partial \beta_i} + \cdots + \theta_q \frac{\partial a_{k-q}}{\partial \beta_i} \\ \qquad \text{当 } \beta_i = \varphi_i, k \geq p+1 \\ a_{k-j} + \theta_1 \frac{\partial a_{k-1}}{\partial \beta_i} + \cdots + \theta_q \frac{\partial a_{k-q}}{\partial \beta_i} \\ \qquad \text{当 } \beta_i = \theta_j,\ k \geq p+1 \end{cases} \tag{3.38}$$

若记 $b_k^{(i)} = \frac{\partial a_k}{\partial \varphi_i}$，则由上式知

$$\theta(B) b_k^{(i)} = 0, \qquad k \geq p+1 \tag{3.39}$$

若记 $d_k^{(j)} = \frac{\partial a_k}{\partial \theta_j}$，则由上式知

$$\theta(B) d_k^{(j)} = a_{k-j}, \qquad k \geq p+1$$

或者进一步有

$$\theta^2(B) d_k^{(j)} = \theta(B) a_{k-j} = \varphi(B) \omega_{k-j}, \qquad k \geq p+1$$

又因取定 $\omega_n = 0$，所以

$$\theta^2(B) d_k^{(j)} = 0, \qquad k \geq p+1+j \tag{3.40}$$

再注意到，ξ 是自由的 $p+q$ 维向量，所以 $\frac{\partial \xi}{\partial \beta_i} = 0$，因此知每个 $\frac{\partial a_k}{\partial \beta_i}$ $(-q+1 \leq k \leq n)$ 都是 ξ 分量的线性组合．这样一来，依定理 1 中对称阵 $X_n^\tau X_n$ 有界的同样论证方法，可以证明 $\frac{\partial X_n^\tau}{\partial \beta_i} \frac{\partial X_n}{\partial \beta_i}$ $(1 \leq i \leq p+q)$ 也有界．从而存在正常数 K，使得对任意的 n，同时成立

$$X_n^\tau X_n \leq K I_{p+q}, \quad \frac{\partial X_n^\tau}{\partial \beta_i} \cdot \frac{\partial X_n}{\partial \beta_i} \leq K I_{p+q}, \quad \pm \frac{\partial Q^{-1}}{\partial \beta_i} \leq K I_{p+q} \tag{3.41}$$

由此还可推得

$$\pm \frac{\partial}{\partial \beta_i}(X_n^\tau X_n) = \pm\left(\frac{\partial X_n^\tau}{\partial \beta_i} X_n + X_n^\tau \frac{\partial X_n}{\partial \beta_i}\right)$$

$$\leqslant X_n^\tau X_n + \frac{\partial X_n^\tau}{\partial \beta_i}\frac{\partial X_n}{\partial \beta_i} \leqslant 2KI_{p+q} \qquad (3.42)$$

最后，利用行列式的微分公式(见[20])：设 A 是依赖于 x 的方阵,则

$$\frac{d}{dx}|A| = |A|\mathrm{tr}\left[A^{-1}\frac{dA}{dx}\right]$$

可知

$$\frac{\partial}{\partial \beta_i}\log|M_n| = \frac{\partial}{\partial \beta_i}[\log|Q^{-1}| - \log|Q^{-1} + X_n^\tau X_n|]$$

$$= |Q|\frac{\partial}{\partial \beta_i}|Q^{-1}| - |Q^{-1} + X_n^\tau X_n|^{-1}\frac{\partial}{\partial \beta_i}|Q^{-1} + X_n^\tau X_n|$$

$$= \mathrm{tr}\left[Q\frac{\partial}{\partial \beta_i}Q^{-1}\right] - \mathrm{tr}\left[(Q^{-1} + X_n^\tau X_n)^{-1}\frac{\partial}{\partial \beta_i}(Q^{-1} + X_n^\tau X_n)\right]$$

于是,由(3.41)和(3.42)式,

$$\left|\frac{\partial}{\partial \beta_i}\log|M_n|\right| \leqslant \left|\mathrm{tr}\left[Q\frac{\partial Q^{-1}}{\partial \beta_i}\right]\right|$$

$$+ \left|\mathrm{tr}\left[(Q^{-1} + X_n^\tau X_n)^{-1}\frac{\partial}{\partial \beta_i}(Q^{-1} + X_n^\tau X_n)\right]\right|$$

$$\leqslant K\mathrm{tr}Q + 3K\mathrm{tr}(Q^{-1} + X_n^\tau X_n)^{-1} \leqslant 4K\mathrm{tr}Q = C_3$$

从而证明了定理 2.

注意，由定理 1 和定理 2 还可推知

$$\left|\frac{\partial|M_n|}{\partial \beta_i}\right| = \left||M_n|\frac{\partial}{\partial \beta_i}\log|M_n|\right| \leqslant \left|\frac{\partial}{\partial \beta_i}\log|M_n|\right| \leqslant C_3 \qquad (3.43)$$

§4. ARMA 序列的信息阵

一、信息阵的定义与 Cramer-Rao 不等式

设 w_t 为平稳随机序列，其分布律依赖于 r 维参数向量 $\boldsymbol{\beta}$，比如正态 ARMA 序列的分布律就是一例. 我们仍以 $\boldsymbol{w}_n$ 表示过程

w_t 的一段样本，并以下列符号分别表示序列的似然函数和对数似然函数

$$L(\boldsymbol{\beta}|\boldsymbol{w}_n)=p(\boldsymbol{w}_n|\boldsymbol{\beta}) \tag{4.1}$$

$$l(\boldsymbol{\beta}|\boldsymbol{w}_n)=\log L(\boldsymbol{\beta}|\boldsymbol{w}_n)=\log p(\boldsymbol{w}_n|\boldsymbol{\beta}) \tag{4.2}$$

其中 $p(\boldsymbol{w}_n|\boldsymbol{\beta})$ 是 $w_1, w_2, \cdots, w_n$ 的联合分布密度函数．在本节讨论中，凡是含义不致混淆的地方，我们用简写记号 $l(\boldsymbol{\beta})$，l 或 $L(\boldsymbol{\beta})$，L 分别表 $l(\boldsymbol{\beta}|\boldsymbol{w}_n)$ 或 $L(\boldsymbol{\beta}|\boldsymbol{w}_n)$．另外，为了明确起见，有时还采用带下标的求均值记号

$$E_{\boldsymbol{\beta}}(f)=\int f(\boldsymbol{w}_n)p(\boldsymbol{w}_n|\boldsymbol{\beta})d\boldsymbol{w}_n$$

表示对参数为 $\boldsymbol{\beta}$ 时的分布密度求 $f(\boldsymbol{w}_n)$ 的均值，而记号 $d\boldsymbol{w}_n$ 是 $dw_1dw_2\cdots dw_n$ 的缩写．对于上述的平稳序列，如果以下的均值存在，

$$J_n(\boldsymbol{\beta})=E_{\boldsymbol{\beta}}\left[\frac{\partial l(\boldsymbol{\beta}|\boldsymbol{w}_n)}{\partial \boldsymbol{\beta}}\frac{\partial l(\boldsymbol{\beta}|\boldsymbol{w}_n)}{\partial \boldsymbol{\beta}^\tau}\right] \tag{4.3}$$

则称 $J_n(\beta)$ 为信息阵，其中

$$\frac{\partial l}{\partial \boldsymbol{\beta}}=\left(\frac{\partial l}{\partial \beta_1},\frac{\partial l}{\partial \beta_2},\cdots,\frac{\partial l}{\partial \beta_r}\right)^\tau$$

$$\frac{\partial l}{\partial \boldsymbol{\beta}^\tau}=\left(\frac{\partial l}{\partial \beta_1},\frac{\partial l}{\partial \beta_2},\cdots,\frac{\partial l}{\partial \beta_r}\right)$$

容易看出，信息阵是非负定的 r 阶方阵．

定理 1. 设平稳序列 w_t 的任何有限维分布都具有分布密度，而且以 $\boldsymbol{\beta}^0$ 为其真实参数．又设基于样本值 $\boldsymbol{w}_n$ 的估计量 $\hat{\boldsymbol{\beta}}=\hat{\boldsymbol{\beta}}(\boldsymbol{w}_n)$ 是 $\boldsymbol{\beta}^0$ 的无偏估计．如果在 $\boldsymbol{\beta}=\boldsymbol{\beta}^0$ 处，以下条件成立：

$$J_n(\boldsymbol{\beta}^0)>0 \tag{4.4}$$

$$E_{\boldsymbol{\beta}}\left(\frac{\partial l(\boldsymbol{\beta}|\boldsymbol{w}_n)}{\partial \boldsymbol{\beta}}\right)_{\boldsymbol{\beta}^0}=0 \tag{4.5}$$

$$E\boldsymbol{\beta}\left[\hat{\boldsymbol{\beta}}(\boldsymbol{w}_n)\left(\frac{\partial l(\boldsymbol{\beta}|\boldsymbol{w}_n)}{\partial \boldsymbol{\beta}^\tau}\right)_{\boldsymbol{\beta}^0}\right]=I_r \tag{4.6}$$

其中记号 $\left(\frac{\partial l}{\partial \boldsymbol{\beta}}\right)_{\boldsymbol{\beta}^0}$ 表示 $\frac{\partial l}{\partial \boldsymbol{\beta}}$ 在 $\boldsymbol{\beta}=\boldsymbol{\beta}_0$ 处的值．那么，

$$\mathrm{var}(\hat{\boldsymbol\beta}(\boldsymbol w_n)) = E_{\beta^0}(\hat{\boldsymbol\beta} - \boldsymbol\beta^0)(\hat{\boldsymbol\beta} - \boldsymbol\beta^0)^\tau \geqslant [J_n(\boldsymbol\beta^0)]^{-1} \quad (4.7)$$

证明. 任取 r 维向量 $\boldsymbol\xi$ 和 $\boldsymbol\zeta$, 由 Schwarz 不等式

$$\left\{E_{\beta^0}\left[\boldsymbol\xi^\tau(\hat{\boldsymbol\beta} - \boldsymbol\beta^0)\left(\frac{\partial l}{\partial \boldsymbol\beta^\tau}\right)_{\beta^0}\boldsymbol\zeta\right]\right\}^2 \leqslant E_{\beta^0}[\boldsymbol\xi^\tau(\hat{\boldsymbol\beta}-\boldsymbol\beta^0)(\hat{\boldsymbol\beta}-\boldsymbol\beta^0)^\tau\boldsymbol\xi]$$

$$\times E_{\beta^0}\left[\boldsymbol\zeta^\tau\left(\frac{\partial l}{\partial \boldsymbol\beta}\right)_{\beta^0}\left(\frac{\partial l}{\partial \boldsymbol\beta^\tau}\right)_{\beta^0}\boldsymbol\zeta\right]$$

再利用 (4.5) 式和 (4.6) 式, 可知上式左端等于 $(\boldsymbol\xi^\tau\boldsymbol\zeta)^2$, 而右端等于 $[\boldsymbol\xi^\tau(\mathrm{var}\hat{\boldsymbol\beta})\boldsymbol\xi]\,[\boldsymbol\zeta^\tau J_n(\boldsymbol\beta^0)\boldsymbol\zeta]$. 由于 $\boldsymbol\xi$ 和 $\boldsymbol\zeta$ 是任意的, 特别取 $\boldsymbol\zeta = [J_n(\boldsymbol\beta^0)]^{-1}\boldsymbol\xi$, 便得到不等式

$$\{\boldsymbol\xi^\tau[J_n(\boldsymbol\beta^0)]^{-1}\boldsymbol\xi\}^2 \leqslant \boldsymbol\xi^\tau(\mathrm{var}\hat{\boldsymbol\beta})\boldsymbol\xi \cdot \boldsymbol\xi^\tau[J_n(\boldsymbol\beta^0)]^{-1}\boldsymbol\xi$$

由此立刻得到 (4.7) 式. 定理得证.

注 1. (4.7) 式称为多维的 Cramér-Rao 不等式.

注 2. 由于

$$E_\beta\frac{\partial l}{\partial \boldsymbol\beta} = \int\frac{1}{L(\boldsymbol\beta|\boldsymbol w_n)}\frac{\partial L(\boldsymbol\beta|\boldsymbol w_n)}{\partial \boldsymbol\beta}p(\boldsymbol w_n|\boldsymbol\beta)d\boldsymbol w_n$$

$$= \int\frac{\partial p(\boldsymbol w_n|\boldsymbol\beta)}{\partial \boldsymbol\beta}d\boldsymbol w_n$$

所以, 若在 $\boldsymbol\beta = \boldsymbol\beta_0$ 处, 上式右端的偏微分号与积分号能够交换, 则条件 (4.5) 式便能成立 (注意 $\int p(\boldsymbol w_n|\boldsymbol\beta)d\boldsymbol w_n = 1$).

注 3. 又由于

$$E_\beta\left[\hat{\boldsymbol\beta}\,\frac{\partial l}{\partial \boldsymbol\beta^\tau}\right] = \int\hat{\boldsymbol\beta}\,\frac{\partial L}{\partial \boldsymbol\beta^\tau}\,\frac{1}{L(\boldsymbol\beta|\boldsymbol w_n)}p(\boldsymbol w_n|\boldsymbol\beta)d\boldsymbol w_n$$

$$= \int\hat{\boldsymbol\beta}\,\frac{\partial p(\boldsymbol w_n|\boldsymbol\beta)}{\partial \boldsymbol\beta^\tau}d\boldsymbol w_n$$

如果在 $\boldsymbol\beta = \boldsymbol\beta^0$ 处上式右端的偏微分号能与积分号交换, 则条件 (4.6) 式便能成立, 因为这时

$$\int\hat{\boldsymbol\beta}\,\frac{\partial p}{\partial \boldsymbol\beta^\tau}d\boldsymbol w_n = \frac{\partial}{\partial \boldsymbol\beta^\tau}\int\hat{\boldsymbol\beta}(\boldsymbol w_n)p(\boldsymbol w_n|\boldsymbol\beta)d\boldsymbol w_n = \frac{\partial}{\partial \boldsymbol\beta^\tau}\boldsymbol\beta = I_r$$

注 4. 如果 $\dfrac{\partial p(\boldsymbol w_n|\boldsymbol\beta)}{\partial \boldsymbol\beta}$ 或 $\hat{\boldsymbol\beta}(\boldsymbol w_n)\dfrac{\partial p(\boldsymbol w_n|\boldsymbol\beta)}{\partial \boldsymbol\beta^\tau}$ 的各分量在 $\boldsymbol\beta^0$ 的某个邻域内对于 $\boldsymbol w_n$ 为一致可积 (一个方便的充分条件是:

它们的绝对值小于一个与$\boldsymbol{\beta}$无关的可积函数)，那么注2或注3中所说的微分和积分运算次序便可以交换，因而(4.5)或(4.6)式成立.

注5. 如果二阶导数$\dfrac{\partial^2 p(\boldsymbol{w}_n|\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\tau}$存在且在$\boldsymbol{\beta}^0$的某个邻域内对于$\boldsymbol{w}_n$为一致可积(同样，一个方便的充分条件是：它的所有元的绝对值小于一个与$\boldsymbol{\beta}$无关的可积函数)，那么对任意的$1\leqslant i$, $j\leqslant r$,

$$E_{\beta}\left[\frac{\partial^2 l(\boldsymbol{\beta}|\boldsymbol{w}_n)}{\partial\beta_i\partial\beta_j}\right]=\int\frac{\partial}{\partial\beta_i}\left[\frac{\partial\log p(\boldsymbol{w}_n|\boldsymbol{\beta})}{\partial\beta_j}\right]p(\boldsymbol{w}_n|\boldsymbol{\beta})d\boldsymbol{w}_n$$

$$=\int\frac{\partial}{\partial\beta_i}\left[\frac{1}{p(\boldsymbol{w}_n|\boldsymbol{\beta})}\cdot\frac{\partial p(\boldsymbol{w}_n|\boldsymbol{\beta})}{\partial\beta_j}\right]p(\boldsymbol{w}_n|\boldsymbol{\beta})d\boldsymbol{w}_n$$

$$=-\int\frac{1}{p^2(\boldsymbol{w}_n|\boldsymbol{\beta})}\frac{\partial p(\boldsymbol{w}_n|\boldsymbol{\beta})}{\partial\beta_i}\frac{\partial p(\boldsymbol{w}_n|\boldsymbol{\beta})}{\partial\beta_j}p(\boldsymbol{w}_n|\boldsymbol{\beta})d\boldsymbol{w}_n$$

$$+\int\frac{\partial^2 p(\boldsymbol{w}_n|\boldsymbol{\beta})}{\partial\beta_i\partial\beta_j}d\boldsymbol{w}_n=-\int\frac{\partial l(\boldsymbol{\beta}|\boldsymbol{w}_n)}{\partial\beta_i}\frac{\partial l(\boldsymbol{\beta}|\boldsymbol{w}_n)}{\partial\beta_j}$$

$$\cdot\, p(\boldsymbol{w}_n|\boldsymbol{\beta})d\boldsymbol{w}_n+\frac{\partial^2}{\partial\beta_i\partial\beta_j}\int p(\boldsymbol{w}_n|\boldsymbol{\beta})d\boldsymbol{w}_n$$

$$=-E_{\beta}\left[\frac{\partial l(\boldsymbol{\beta}|\boldsymbol{w}_n)}{\partial\beta_i}\frac{\partial l(\boldsymbol{\beta}|\boldsymbol{w}_n)}{\partial\beta_j}\right]$$

于是我们得到

$$J_n(\boldsymbol{\beta})=-E_{\beta}\left[\frac{\partial^2 l(\boldsymbol{\beta}|\boldsymbol{w}_n)}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\tau}\right]\tag{4.8}$$

根据定理1和各条注解，我们来讨论正态 ARMA 序列的信息阵的一般性质.

设序列模型为

$$\varphi(B)w_t=\theta(B)a_t$$

则其协方差函数为

$$\gamma_k=\sigma_a^2\int_{-\frac{1}{2}}^{\frac{1}{2}}e^{2\pi ik\mu}\frac{|\theta(e^{-2\pi i\mu})|^2}{|\varphi(e^{-2\pi i\mu})|^2}d\mu$$

$$=\frac{\sigma_a^2}{2\pi i}\oint_{|z|=1}\frac{z^{k-1}\theta(z)\theta(z^{-1})}{\varphi(z)\varphi(z^{-1})}dz$$

γ_k 是参数 $\boldsymbol{\varphi}$ 和 $\boldsymbol{\theta}$ 的连续函数．由于 $\varphi(z)$ 的根都在单位圆外，易知这时前面注 4 和注 5 中所说的积分微分换序的充分条件成立，所以有

$$\frac{\partial \gamma_k}{\partial \varphi_j}=\frac{\sigma_a^2}{2\pi i}\oint_{|z|=1}\frac{z^{k-1+j}\varphi(z^{-1})+z^{k-1-j}\varphi(z)}{[\varphi(z)\varphi(z^{-1})]^2}\theta(z)\theta(z^{-1})dz$$

$$\frac{\partial \gamma_k}{\partial \theta_j}=\frac{-\sigma_a^2}{2\pi i}\oint_{|z|=1}\frac{z^{k-1+j}\theta(z^{-1})+z^{k-1-j}\theta(z)}{\varphi(z)\varphi(z^{-1})}dz$$

由此可见，在 $\boldsymbol{\beta}$ 的平稳域内，γ_k 的一阶偏导数仍是 $\boldsymbol{\beta}$ 的连续函数，类似地可以证明，γ_k 的其它各阶导数也存在而且关于 $\boldsymbol{\beta}$ 为连续．由附录 §3，正态 ARMA 序列的分布密度函数和对数似然函数分别为（见附录 §3 (3.18)，(3.19) 式）

$$p(\boldsymbol{w}_n|\boldsymbol{\beta},\sigma_a^2)=(2\pi\sigma_a^2)^{-n/2}|M_n|^{\frac{1}{2}}\exp\left\{-\frac{S_n(\boldsymbol{\beta})}{2\sigma_a^2}\right\}$$

$$l(\boldsymbol{\beta},\sigma_a^2|\boldsymbol{w}_n)=-\frac{n}{2}\log 2\pi-\frac{1}{2}\left\{n\log\sigma_a^2-\log|M_n|+\frac{1}{\sigma_a^2}S_n(\boldsymbol{\beta})\right\} \tag{4.9}$$

其中 $M_n=\sigma_a^2(E\boldsymbol{w}_n\boldsymbol{w}_n^{\tau})^{-1}=\sigma_a^2\Gamma_n^{-1}$．

用矩阵微分公式（见[20]）可知

$$\frac{\partial M_n}{\partial \beta_j}=-M_n\frac{\partial M_n^{-1}}{\partial \beta_j}M_n=-\sigma_a^{-2}M_n\frac{\partial \Gamma_n}{\partial \beta_j}M_n \tag{4.10}$$

$$\frac{\partial}{\partial \beta_j}S_n(\boldsymbol{\beta})=\frac{\partial}{\partial \beta_j}(\boldsymbol{w}_n^{\tau}M_n\boldsymbol{w}_n)=-\sigma_a^{-2}\boldsymbol{w}_n^{\tau}M_n\frac{\partial \Gamma_n}{\partial \beta_j}M_n\boldsymbol{w}_n \tag{4.11}$$

$$\frac{\partial}{\partial \beta_j}\log|M_n|=\mathrm{tr}\left(M_n^{-1}\frac{\partial M_n}{\partial \beta_j}\right)=-\mathrm{tr}\left(\frac{\partial M_n^{-1}}{\partial \beta_j}M_n\right)=-\sigma_a^{-2}\mathrm{tr}\left(\frac{\partial \Gamma_n}{\partial \beta_j}M_n\right) \tag{4.12}$$

其中 $\frac{\partial \Gamma_n}{\partial \beta_j}$ 的元都是 $\gamma_k(0\leqslant k\leqslant n-1)$ 关于 β_j 的偏导数，而前面已证明了 $\frac{\partial \gamma_k}{\partial \beta_j}$ 存在且连续，所以以上三式的偏导数都存在且连

续．又由于 $\boldsymbol{w}_n$ 是正态分布的，因此，$E_{\boldsymbol{\beta}} \frac{\partial S_n(\boldsymbol{\beta})}{\partial \beta_i}$ 和 $E_{\boldsymbol{\beta}} \frac{\partial S_n(\boldsymbol{\beta})}{\partial \beta_i} \cdot \frac{\partial S_n(\boldsymbol{\beta})}{\partial \beta_j}$ 必存在．于是由(4.9)式易见，对于正态 ARMA 序列而言，$J_n(\boldsymbol{\beta}) = E_{\boldsymbol{\beta}}\left(\frac{\partial l}{\partial \boldsymbol{\beta}} \frac{\partial l}{\partial \boldsymbol{\beta}^\tau}\right)$ 必存在．

另外，容易求出

$$\frac{\partial p(\boldsymbol{w}_n|\boldsymbol{\beta},\sigma_a^2)}{\partial \beta_i} = \frac{1}{2}\left(|M_n|^{-1} \frac{\partial |M_n|}{\partial \beta_i} - \frac{1}{\sigma_a^2} \frac{\partial S_n(\boldsymbol{\beta})}{\partial \beta_i}\right) \cdot p(\boldsymbol{w}_n|\boldsymbol{\beta},\sigma_a^2)$$

$$\begin{aligned}\frac{\partial^2 p(\boldsymbol{w}_n|\boldsymbol{\beta},\sigma_a^2)}{\partial \beta_i \partial \beta_j} = \frac{1}{2}\Big\{&|M_n|^{-1} \frac{\partial^2 |M_n|}{\partial \beta_i \partial \beta_j} - \frac{1}{\sigma_a^2} \frac{\partial^2 S_n(\boldsymbol{\beta})}{\partial \beta_i \partial \beta_j} + \frac{1}{2} \\ &\cdot \left(|M_n|^{-2} \frac{\partial |M_n|}{\partial \beta_i} \frac{\partial |M_n|}{\partial \beta_j} + \frac{1}{\sigma_a^4} \frac{\partial S_n(\boldsymbol{\beta})}{\partial \beta_i}\right. \\ &\left.\cdot \frac{\partial S_n(\boldsymbol{\beta})}{\partial \beta_j}\right) - \frac{|M_n|^{-1}}{2\sigma_a^2}\left(\frac{\partial |M_n|}{\partial \beta_i} \frac{\partial S_n(\boldsymbol{\beta})}{\partial \beta_j}\right. \\ &\left.+ \frac{\partial S_n(\boldsymbol{\beta})}{\partial \beta_i} \frac{\partial |M_n|}{\partial \beta_j}\right)\Big\} p(\boldsymbol{w}_n|\boldsymbol{\beta},\sigma_a^2)\end{aligned}$$

由于 γ_k 是 $\boldsymbol{\beta}$ 的连续函数，而在附录§3中又已证明了$1 \geqslant |M_n| > 0$，所以在 $\boldsymbol{\beta}^0$ 的一个适当的邻域内必满足 $M_n \geqslant \varepsilon I_n$，于是有

$$S_n(\boldsymbol{\beta}) = \boldsymbol{w}_n^\tau M_n \boldsymbol{w}_n \geqslant \varepsilon \boldsymbol{w}_n^\tau \boldsymbol{w}_n$$

又由于 $\frac{\partial \gamma_k}{\partial \beta_i}$，$\frac{\partial^2 \gamma_k}{\partial \beta_i \partial \beta_j}$ $(1 \leqslant i, j \leqslant r)$ 都是 $\boldsymbol{\beta}$ 的连续函数，从而 $\frac{\partial M_n}{\partial \beta_i}$，$\frac{\partial^2 M_n}{\partial \beta_i \partial \beta_j}$ 也都连续(见(4.10)式)，故在上述邻域内都有界．因此存在正常数 K_n，使当 $\boldsymbol{\beta}$ 在邻域内时，

$$\left|\frac{\partial S_n(\boldsymbol{\beta})}{\partial \beta_i}\right| \leqslant K_n \boldsymbol{w}_n^\tau \boldsymbol{w}_n, \quad \left|\frac{\partial^2 S_n(\boldsymbol{\beta})}{\partial \beta_i \partial \beta_j}\right| \leqslant K_n \boldsymbol{w}_n^\tau \boldsymbol{w}_n$$

但此时

$$K_n \boldsymbol{w}_n^\tau \boldsymbol{w}_n p(\boldsymbol{w}_n|\boldsymbol{\beta},\sigma_a^2) \leqslant K_n \boldsymbol{w}_n^\tau \boldsymbol{w}_n (2\pi\sigma_a^2)^{-n/2} e^{-\frac{\varepsilon}{2\sigma_a^2} \boldsymbol{w}_n^\tau \boldsymbol{w}_n}$$

$$K_n^2(\boldsymbol{w}_n^\tau \boldsymbol{w}_n)^2 p(\boldsymbol{w}_n | \boldsymbol{\beta}, \sigma_a^2) \leqslant K_n^2(\boldsymbol{w}_n^\tau \boldsymbol{w}_n)^2 (2\pi\sigma_a^2)^{-n/2} e^{-\frac{\varepsilon}{2\sigma_a^2}\boldsymbol{w}_n^\tau \boldsymbol{w}_n}$$

上两式右边都是与$\boldsymbol{\beta}$无关的$\boldsymbol{w}_n$的可积函数，可见两式左边当$\boldsymbol{\beta}$在$\boldsymbol{\beta}_0$的某个邻域内时是一致可积的．综上所述，易知$\dfrac{\partial p(\boldsymbol{w}_n | \boldsymbol{\beta}, \sigma_a^2)}{\partial \beta_i}$，$\dfrac{\partial^2 p(\boldsymbol{w}_n | \boldsymbol{\beta}, \sigma_a^2)}{\partial \beta_i \partial \beta_j}$作为$\boldsymbol{w}_n$的函数关于$\boldsymbol{\beta}$为一致可积．于是由定理1的注4和注5，对于正态 ARMA 序列，(4.5)式成立，且其信息矩阵可由(4.8)式表出．用上述类似的推论，还可以证明$J_n(\beta)$的连续性和可微性．

二、正态 AR 序列的大样本信息阵

按照以前的记号，记正态 AR 序列的参数向量为$\boldsymbol{\alpha}^\tau = (\boldsymbol{\varphi}^\tau, \sigma_a^2)$，其分布函数为

$$p(\boldsymbol{w}_n | \boldsymbol{\varphi}, \sigma_a^2) = p(\boldsymbol{w}_n | \boldsymbol{\alpha}) = (2\pi\sigma_a^2)^{-n/2} |M_p|^{\frac{1}{2}} \exp\left\{-\frac{S_n(\boldsymbol{\varphi})}{2\sigma_a^2}\right\}$$

$$l(\boldsymbol{\alpha} | \boldsymbol{w}_n) = \log p(\boldsymbol{w}_n | \boldsymbol{\alpha}) = -\frac{1}{2}\left\{n \log 2\pi + n \log \sigma_a^2 - \log |M_p| + \frac{S_n(\boldsymbol{\varphi})}{\sigma_a^2}\right\}$$

其中 $M_p = \sigma_a^2 \Gamma_p^{-1} = \sigma_a^2 (E\boldsymbol{w}_p \boldsymbol{w}_p^\tau)^{-1}$．依本附录§3可知，$M_p$的诸元是$\boldsymbol{\varphi}$的二次型函数，且当 $n \geqslant 2p$ 时有

$$S_n(\boldsymbol{\varphi}) = \sum_{i,j=0}^{p} d_{ij}\varphi_i\varphi_j, \qquad (\text{其中 } \varphi_0 = -1) \tag{4.13}$$

$$d_{ij} = d_{ji} = \sum_{t=1+j}^{n-i} w_{t+i-j} w_t, \quad (0 \leqslant i, j \leqslant p) \tag{4.14}$$

由此可得

$$\frac{\partial l}{\partial \sigma_a^2} = -\frac{n}{2\sigma_a^2} + \frac{S_n(\boldsymbol{\varphi})}{2\sigma_a^4} \tag{4.15a}$$

$$\frac{\partial l}{\partial \varphi_i} = \frac{1}{2}\frac{\partial}{\partial \varphi_i} \log |M_p| - \frac{1}{\sigma_a^2}\sum_{j=0}^{p} d_{ij}\varphi_j, \quad (1 \leqslant i \leqslant p) \tag{4.15b}$$

$$-\frac{\partial^2 l}{\partial\sigma_a^2\partial\sigma_a^2} = -\frac{n}{2\sigma_a^4} + \frac{S_n(\varphi)}{\sigma_a^6} \tag{4.15c}$$

$$-\frac{\partial^2 l}{\partial\varphi_i\partial\varphi_j} = -\frac{1}{2}\frac{\partial^2}{\partial\varphi_i\partial\varphi_j}\log|M_p| + \frac{1}{\sigma_a^2}d_{ij}, \quad (1 \leqslant i, j \leqslant p) \tag{4.15d}$$

$$-\frac{\partial^2 l}{\partial\sigma_a^2\partial\varphi_i} = -\frac{1}{\sigma_a^4}\sum_{j=0}^{p} d_{ij}\varphi_j = \frac{1}{\sigma_a^2}\left(\frac{\partial l}{\partial\varphi_i} - \frac{1}{2}\frac{\partial}{\partial\varphi_i}\log|M_p|\right)$$

$$(1 \leqslant i \leqslant p) \tag{4.15e}$$

依前一小节的论述，对于正态 ARMA 序列，(4.5) 式和 (4.8) 式都成立，依此及 (4.15a) 式可得

$$E_a S_n(\varphi) = n\sigma_a^2 \tag{4.16}$$

依 (4.14) 式和 (4.15b) 又得 (注意 $\sum_{j=0}^{p}\varphi_j\gamma_{i-j} = 0, \ i \geqslant 1$)

$$E_a d_{ij} = (n - i - j)\gamma_{i-j} \qquad (1 \leqslant i, j \leqslant p) \tag{4.17a}$$

$$\frac{\partial}{\partial\varphi_i}\log|M_p| = 2\sigma_a^{-2}\sum_{j=0}^{p}(n - i - j)\varphi_j\gamma_{i-j}$$

$$= -2\sigma_a^{-2}\sum_{j=1}^{p} j\varphi_j\gamma_{i-j} \tag{4.17b}$$

再由 (4.15c), (4.15d), 和 (4.15e) 可得

$$-E_a\frac{\partial^2 l}{\partial\sigma_a^2\partial\sigma_a^2} = \frac{n}{2\sigma_a^4} \tag{4.18a}$$

$$-E_a\frac{\partial^2 l}{\partial\varphi_i\partial\varphi_j} = -\frac{1}{2}\frac{\partial^2}{\partial\varphi_i\partial\varphi_j}\log|M_p| + \frac{n-i-j}{\sigma_a^2}\gamma_{i-j}$$

$$(1 \leqslant i, j \leqslant p) \tag{4.18b}$$

$$-E_a\frac{\partial^2 l}{\partial\sigma_a^2\partial\varphi_i} = -\frac{1}{2\sigma_a^2}\frac{\partial}{\partial\varphi_i}\log|M_p| = \frac{1}{\sigma_a^4}\sum_{j=1}^{p} j\varphi_j\gamma_{i-j}$$

$$(1 \leqslant i \leqslant p) \tag{4.18c}$$

以上三式给出了 $J_n(\alpha)$ 诸元的明确公式．当 n 很大时，$J_n(\alpha)$ 的渐近公式也可由它们得到，即

$$J_n(\alpha)=\frac{n}{\sigma_a^2}\begin{pmatrix}\Gamma_p & 0\\ 0 & \frac{1}{2\sigma_a^2}\end{pmatrix}+O(1) \tag{4.19}$$

这里的 $O(1)$ 表示与 n 无关的 $p+1$ 阶常值矩阵. 由此式容易看出，当 n 很大以后，

$$J_n^{-1}(\alpha)=\frac{\sigma_a^2}{n}\begin{pmatrix}\Gamma_p^{-1} & 0\\ 0 & 2\sigma_a^2\end{pmatrix}+o\left(\frac{1}{n}\right)=\frac{1}{n}\begin{pmatrix}M_p & 0\\ 0 & 2\sigma_a^4\end{pmatrix}+o\left(\frac{1}{n}\right) \tag{4.20}$$

其中 $o\left(\frac{1}{n}\right)$ 表示一个 $p+1$ 阶矩阵，当 $n\to\infty$ 时，它的诸元是比 $\frac{1}{n}$ 更高阶的无穷小量. M_p 的表达式由本附录 §3 的 (3.15) 式给出.

例如，AR(1，0) 序列的信息阵之逆为

$$J_n^{-1}(\alpha)=\frac{1}{n}\begin{pmatrix}1-\varphi_1^2 & 0\\ 0 & 2\sigma_a^4\end{pmatrix}+o\left(\frac{1}{n}\right) \tag{4.21}$$

AR(2，0) 序列的信息阵之逆为

$$J_n^{-1}(\alpha)=\frac{1}{n}\begin{pmatrix}1-\varphi_2^2 & -\varphi_1(1+\varphi_2) & 0\\ -\varphi_1(1+\varphi_2) & 1-\varphi_2^2 & 0\\ 0 & 0 & 2\sigma_a^4\end{pmatrix}+o\left(\frac{1}{n}\right) \tag{4.22}$$

三、正态 ARMA 序列的大样本信息阵

关于一般正态 ARMA(p，q) 序列 w_t 的信息阵，其表达式没有 AR 序列那样简单. 在下面的讨论中，我们暂时先把 σ_a^2 当做已知量，而设 w_t 的分布律依赖于某参数向量 $\boldsymbol{\beta}$. 由本附录 §3 的 (3.30) 式，$\boldsymbol{\beta}$ 的对数似然函数为

$$l(\boldsymbol{\beta}\mid \boldsymbol{w}_n)=-\frac{1}{2}\left\{n\log 2\pi+n\log\sigma_a^2-f_n(\boldsymbol{\beta})+\frac{1}{\sigma_a^2}S_n(\boldsymbol{\beta})\right\} \tag{4.23}$$

其中

$$f_n(\boldsymbol{\beta}) = \log|M_n|, \qquad S_n(\boldsymbol{\beta}) = \boldsymbol{w}_n^\tau M_n \boldsymbol{w}_n$$

用 $\mu_{kl}^{(n)}(1 \leqslant k, l \leqslant n)$ 表 $M_n = \sigma_a^2 \Gamma_n^{-1}$ 的第 k 行 l 列元．由微分公式(4.12)可知

$$E_\beta\left(\frac{\partial S_n}{\partial \beta_i}\right) = E_\beta\left(\boldsymbol{w}_n^\tau \frac{\partial M_n}{\partial \beta_i} \boldsymbol{w}_n\right) = \sum_{k,l=1}^{n} \frac{\partial \mu_{kl}^{(n)}}{\partial \beta_i} \gamma_{k-l}$$

$$= \operatorname{tr}\left(\frac{\partial M_n}{\partial \beta_i} \Gamma_n\right) = \sigma_a^2 \operatorname{tr}\left(\frac{\partial M_n}{\partial \beta_i} M_n^{-1}\right)$$

$$= \sigma_a^2 \frac{\partial}{\partial \beta_i} \log|M_n| = \sigma_a^2 \frac{\partial f_n}{\partial \beta_i} \tag{4.24}$$

又由本附录(2.5)式和逆矩阵微分公式(4.10)知

$$E_\beta\left(\frac{\partial S_n}{\partial \beta_i} \frac{\partial S_n}{\partial \beta_j}\right) = E_\beta\left(\boldsymbol{w}_n^\tau \frac{\partial M_n}{\partial \beta_i} \boldsymbol{w}_n\right)\left(\boldsymbol{w}_n^\tau \frac{\partial M_n}{\partial \beta_j} \boldsymbol{w}_n\right)$$

$$= \sum_{k,l,s,t=1}^{n} \frac{\partial \mu_{kl}^{(n)}}{\partial \beta_i} \frac{\partial \mu_{st}^{(n)}}{\partial \beta_j} E_\beta w_k w_l w_s w_t$$

$$= \sum_{k,l,s,t=1}^{n} \frac{\partial \mu_{kl}^{(n)}}{\partial \beta_i} \frac{\partial \mu_{st}^{(n)}}{\partial \beta_j} (\gamma_{k-l}\gamma_{s-t} + \gamma_{k-s}\gamma_{l-t} + \gamma_{k-t}\gamma_{l-s})$$

$$= \operatorname{tr}\left(\frac{\partial M_n}{\partial \beta_i} \Gamma_n\right) \operatorname{tr}\left(\frac{\partial M_n}{\partial \beta_j} \Gamma_n\right) + 2\operatorname{tr}\left(\frac{\partial M_n}{\partial \beta_i} \Gamma_n \frac{\partial M_n}{\partial \beta_j} \Gamma_n\right)$$

$$= \sigma_a^4 \frac{\partial f_n}{\partial \beta_i} \frac{\partial f_n}{\partial \beta_j} + 2\operatorname{tr}\left(M_n \frac{\partial \Gamma_n}{\partial \beta_i} M_n \frac{\partial \Gamma_n}{\partial \beta_j}\right) \tag{4.25}$$

于是由(4.23)，(4.24)，(4.25)三式得

$$E_\beta\left(\frac{\partial l}{\partial \beta_i} \frac{\partial l}{\partial \beta_j}\right) = \frac{1}{4} E_\beta\left(\frac{1}{\sigma_a^2} \frac{\partial S_n}{\partial \beta_i} - \frac{\partial f_n}{\partial \beta_i}\right)$$

$$\cdot\left(\frac{1}{\sigma_a^2} \frac{\partial S_n}{\partial \beta_j} - \frac{\partial f_n}{\partial \beta_j}\right) = \frac{1}{4}\left\{\frac{1}{\sigma_a^4} E_\beta\left(\frac{\partial S_n}{\partial \beta_i} \frac{\partial S_n}{\partial \beta_j}\right)\right.$$

$$\left. - \frac{\partial f_n}{\partial \beta_i} \frac{\partial f_n}{\partial \beta_j}\right\} = \frac{1}{2\sigma_a^4} \operatorname{tr}\left(M_n \frac{\partial \Gamma_n}{\partial \beta_i} M_n \frac{\partial \Gamma_n}{\partial \beta_j}\right) \tag{4.26}$$

现在，具体取参数 $\boldsymbol{\beta}^\tau = (\boldsymbol{\varphi}^\tau, \boldsymbol{\theta}^\tau)$．为进一步估算(4.26)式，需要计算 $\frac{\partial \Gamma_n}{\partial \beta_i}(1 \leqslant i \leqslant p+q)$．注意，若表

$$\frac{1}{\varphi(z)}=\sum_{u=0}^{\infty}\varphi^{(u)}z^{u},\quad \frac{1}{\theta(z)}=\sum_{u=0}^{\infty}\theta^{(u)}z^{u} \tag{4.27}$$

则由前面(4.8)式到(4.9)式之间的推导，

$$\begin{aligned}\frac{\partial\gamma_t}{\partial\varphi_j}&=\frac{\sigma_a^2}{2\pi i}\oint_{|z|=1}\frac{z^{t-1+j}\varphi(z^{-1})+z^{t-1-j}\varphi(z)}{(\varphi(z)\varphi(z^{-1}))^2}\theta(z)\theta(z^{-1})dz\\&=\frac{\sigma_a^2}{2\pi i}\oint_{|z|=1}\left(\frac{z^{t-1+j}}{\varphi(z)}+\frac{z^{t-1-j}}{\varphi(z^{-1})}\right)\frac{\theta(z)\theta(z^{-1})}{\varphi(z)\varphi(z^{-1})}dz\\&=\frac{\sigma_a^2}{2\pi i}\oint_{|z|=1}\sum_{u=0}^{\infty}\varphi^{(u)}(z^{t+j+u-1}+z^{t-j-u-1})\frac{\theta(z)\theta(z^{-1})}{\varphi(z)\varphi(z^{-1})}dz\\&=\sum_{u=0}^{\infty}\varphi^{(u)}(\gamma_{t+j+u}+\gamma_{t-j-u}),\quad 1\leqslant j\leqslant p\end{aligned} \tag{4.28}$$

$$\begin{aligned}\frac{\partial\gamma_t}{\partial\theta_j}&=-\frac{\sigma_a^2}{2\pi i}\oint_{|z|=1}\frac{z^{t-1+j}\theta(z^{-1})+z^{t-1-j}\theta(z)}{\varphi(z)\varphi(z^{-1})}dz\\&=-\frac{\sigma_a^2}{2\pi i}\oint_{|z|=1}\left(\frac{z^{t-1+j}}{\theta(z)}+\frac{z^{t-1-j}}{\theta(z^{-1})}\right)\frac{\theta(z)\theta(z^{-1})}{\varphi(z)\varphi(z^{-1})}dz\\&=-\sum_{u=0}^{\infty}\theta^{(u)}(\gamma_{t+j+u}+\gamma_{t-j-u}),\quad 1\leqslant j\leqslant q\end{aligned} \tag{4.29}$$

将(4.28)式代入(4.26)式的右边，即得，对 $1\leqslant i,\ j\leqslant p$，

$$\begin{aligned}E_\beta\left(\frac{\partial l}{\partial\varphi_i}\frac{\partial l}{\partial\varphi_j}\right)&=\frac{1}{2\sigma_a^4}\sum_{t_1,t_2,s_1,s_2=1}^{n}\mu_{t_1t_2}^{(n)}\frac{\partial\gamma_{t_1-s_1}}{\partial\varphi_i}\mu_{s_1s_2}^{(n)}\frac{\partial\gamma_{t_2-s_2}}{\partial\varphi_j}\\&=\frac{1}{2\sigma_a^4}\sum_{t_1,t_2,s_1,s_2=1}^{n}\mu_{t_1t_2}^{(n)}\left\{\sum_{u=0}^{\infty}\varphi^{(u)}(\gamma_{t_1-s_1+i+u}+\gamma_{t_1-s_1-i-u})\right\}\mu_{s_1s_2}^{(n)}\\&\quad\cdot\left\{\sum_{v=0}^{\infty}\varphi^{(v)}(\gamma_{t_2-s_2+j+v}+\gamma_{t_2-s_2-j-v})\right\}\\&=\frac{1}{2\sigma_a^4}\sum_{u,v=0}^{\infty}\varphi^{(u)}\varphi^{(v)}\sum_{t_1,t_2,s_1,s_2=1}^{n}\mu_{t_1t_2}^{(n)}\mu_{s_1s_2}^{(n)}(\gamma_{t_1-s_1+i+u}\gamma_{t_2-s_2+j+v}\\&\quad+\gamma_{t_1-s_1+i+u}\gamma_{t_2-s_2-j-v}+\gamma_{t_1-s_1-i-u}\gamma_{t_2-s_2+j+v}\\&\quad+\gamma_{t_1-s_1-i-u}\gamma_{t_2-s_2-j-v})\end{aligned} \tag{4.30}$$

我们分别计算上式中的四项求和. 以第一项为例，并且又将它分成以下四个部分：

$$\sum_{u,v=0}^{\infty}\varphi^{(u)}\varphi^{(v)}\sum_{t_1,t_2,s_1,s_2=1}^{n}\mu_{t_1t_2}^{(n)}\mu_{s_1s_2}^{(n)}\gamma_{s_1-t_1+i+u}\gamma_{t_2-s_2+j+v}$$

$$=\sum_{u,v=0}^{\infty}\varphi^{(u)}\varphi^{(v)}\Big(\sum_{\substack{t_1+i+u\leqslant n\\ s_2-j-v\geqslant 1}}+\sum_{\substack{t_1+i+u\leqslant n\\ s_2-j-v\leqslant 0}}+\sum_{\substack{t_1+i+u\geqslant n+1\\ s_2-j-v\geqslant 1}}$$

$$+\sum_{\substack{t_1+i+u\geqslant n+1\\ s_2-j-v\leqslant 0}}\Big)\sum_{t_2=1}^{n}\mu_{t_1t_2}^{(n)}\gamma_{t_2-(s_2-j-v)}\sum_{s_1=1}^{n}\gamma_{t_1+i+u-s_1}\mu_{s_1s_2}^{(n)}$$

$$=S_1+S_2+S_3+S_4 \tag{4.31}$$

在对它们进行估算时，还须用到几个简单的事实．首先由于 $M_n=\sigma_a^2\Gamma_n^{-1}$，所以当 $1\leqslant s,t\leqslant n$ 时，

$$\sum_{t_2=1}^{n}\mu_{t_1t_2}^{(n)}\gamma_{t_2-s}=\sigma_a^2\delta_{t_1s},\quad \sum_{s_1=1}^{n}\gamma_{t-s_1}\mu_{s_1s_2}^{(n)}=\sigma_a^2\delta_{ts_2} \tag{4.32}$$

其次，由附录 §3 的 (3.32) 式

$$0<\sigma_a^{2n}\leqslant|\Gamma_n|=\sigma_a^{2n}|M_n|^{-1}\leqslant\frac{\sigma_a^{2n}}{c_1}<\infty$$

其中 c_1 为与 n 无关的常数．若令 Γ_n^{kk} 表 Γ_n 的 k 行 k 列元的代数余子式，则显然可表

$$\Gamma_n^{kk}=\begin{vmatrix}\Gamma_{k-1} & H\\ H^\tau & \Gamma_{n-k}\end{vmatrix}$$

其中 H 不必具体写出．于是由分块阵的行列式公式和附录 §3 中的引理，

$$\mu_{kk}^{(n)}=\frac{\Gamma_n^{kk}}{|\Gamma_n|}\sigma_a^2=\frac{|\Gamma_{k-1}|\,|\Gamma_{n-k}-H^\tau\Gamma_{k-1}^{-1}H|}{|\Gamma_n|}\sigma_a^2$$

$$\leqslant\frac{|\Gamma_{k-1}|\,|\Gamma_{n-k}|}{|\Gamma_n|}\sigma_a^2\leqslant\frac{1}{\sigma_a^{2n}}\cdot\frac{\sigma_a^{2(k-1)}}{c_1}\cdot\frac{\sigma_a^{2(n-k)}}{c_1}\sigma_a^2$$

$$=\frac{1}{c_1^2}<\infty$$

又由 M_n 的正定性知，对任何 $1\leqslant k,l\leqslant n$，

$$|\mu_{kl}^{(n)}|\leqslant\sqrt{\mu_{kk}^{(n)}\mu_{ll}^{(n)}}\leqslant\frac{1}{c_1^2}<\infty \tag{4.33}$$

再次，由于 $\varphi(z)$，$\theta(z)$ 的根都在单位圆外，所以 $|\varphi^{(u)}|$ 和 $|\theta^{(u)}|$ 都是依 u 的负指数律下降的，因而

$$
\sum_{u=0}^{\infty} |\varphi^{(u)}| < \infty, \quad \sum_{u=0}^{\infty} |\theta^{(u)}| < \infty, \quad \sum_{u=0}^{\infty} u|\varphi^{(u)}| < \infty,
$$

$$
\sum_{u=0}^{\infty} u|\theta^{(u)}| < \infty \tag{4.34}
$$

最后，由于 $|\gamma_t|$ 也依 t 的负指数律下降，所以有

$$
\gamma = \sum_{k=0}^{\infty} |\gamma_k| < \infty \tag{4.35}
$$

根据 (4.32)—(4.35) 诸式，我们可以对 (4.31) 式中的 S_1, S_2, S_3, S_4 分别作以下的估算:

$$
\begin{aligned}
S_1 &= \sum_{u,v=0}^{\infty} \varphi^{(u)}\varphi^{(v)} \sum_{\substack{t_1+i+u\leqslant n \\ t_2-j-v\geqslant 1}} \sigma_a^4 \delta_{t_1,t_2-j-v}\delta_{t_1+i+u,t_2} \\
&= \sigma_a^4 \sum_{u=\max(0,j-i)}^{n-i} \varphi^{(u)}\varphi^{(u+i-j)}(n-u-i) \\
&= n\sigma_a^4 \sum_{u=\max(i,j)}^{\infty} \varphi^{(u-i)}\varphi^{(u-j)} + O(1)
\end{aligned} \tag{4.36}
$$

$$
\begin{aligned}
|S_2| &= \sigma_a^2 \left| \sum_{u,v=0}^{\infty} \varphi^{(u)}\varphi^{(v)} \sum_{\substack{t_1+i+u\leqslant n \\ t_2-j-v\leqslant 0}} \left(\sum_{s_2=1}^{n} \mu_{t_1 s_2}^{(n)} \gamma_{s_2-(t_2-j-v)} \right) \delta_{t_1+i+u,t_2} \right| \\
&\leqslant \frac{\gamma\sigma_a^2}{c_1^2} \sum_{u=0}^{\infty} |\varphi^{(u)}| \sum_{v=0}^{\infty} (v+j)|\varphi^{(v)}| \\
&= O(1)
\end{aligned} \tag{4.37}
$$

$$
\begin{aligned}
|S_3| &= \sigma_a^2 \left| \sum_{u,v=0}^{\infty} \varphi^{(u)}\varphi^{(v)} \sum_{\substack{t_1+i+u\geqslant n+1 \\ t_2-j-v\geqslant 1}} \delta_{t_1,t_2-j-v} \left(\sum_{s_1=1}^{n} \gamma_{t_1+i+u-s_1} \mu_{s_1 t_2}^{(n)} \right) \right| \\
&\leqslant \frac{\gamma\sigma_a^2}{c_1^2} \sum_{u=0}^{\infty} (u+i)|\varphi^{(u)}| \sum_{v=0}^{\infty} |\varphi^{(v)}| \\
&= O(1)
\end{aligned} \tag{4.38}
$$

$$|S_4| \leqslant \frac{\gamma^2\sigma_a^2}{c_1^4}\sum_{u=0}^{\infty}(u+i)|\varphi^{(u)}|\sum_{v=0}^{\infty}(v+j)|\varphi^{(v)}| = O(1) \tag{4.39}$$

联合以上四式即有

$$\sum_{u,v=0}^{\infty}\varphi^{(u)}\varphi^{(v)}\sum_{t_1,t_2,s_1,s_2=1}^{n}\mu_{t_1t_2}^{(n)}\mu_{s_1s_2}^{(n)}\gamma_{t_1-s_1+i-u}\gamma_{t_2-s_2+j+v}$$

$$= n\sigma_a^4\sum_{u=\max(i,j)}^{\infty}\varphi^{(u-i)}\varphi^{(u-j)} + O(1) \tag{4.40}$$

用类似的方法可以估算(4.30)式中其余三项求和的值；在此三项所对应的(4.31)式的分项求和中，S_2，S_3，S_4 仍为 $O(1)$ 阶，所不同的只是，它们对应于(4.31)式的第一项 S_1 的估计与(4.36)式不一样，这三项求和依次为

$$\sigma_a^4\sum_{u,v=0}^{\infty}\varphi^{(u)}\varphi^{(v)}\sum_{\substack{t_1+i+u\leqslant n\\ s_2+j+v\leqslant n}}\delta_{t_2,s_2+j+v}\,\delta_{t_1+i+u,s_2} = 0$$

$$\sigma_a^4\sum_{u,v=0}^{\infty}\varphi^{(u)}\varphi^{(v)}\sum_{\substack{t_1-i-u\geqslant 1\\ s_2-j-v\geqslant 1}}\delta_{t_1,s_2-j-v}\,\delta_{t_1-i-u,t_2} = 0$$

$$\sigma_a^4\sum_{u,v=0}^{\infty}\varphi^{(u)}\varphi^{(v)}\sum_{\substack{t_1-i-u\geqslant 1\\ s_2+j+v\leqslant n}}\delta_{t_2,s_2+j+v}\,\delta_{t_1-i-u,s_2}$$

$$= n\sigma_a^4\sum_{u=\max(i,j)}^{\infty}\varphi^{(u-i)}\varphi^{(u-j)} + O(1)$$

将这些结果与(4.40)式一道代入(4.30)式，即得

$$E_\beta\left(\frac{\partial l}{\partial\varphi_i}\frac{\partial l}{\partial\varphi_j}\right) = n\sum_{u=\max(i,j)}^{\infty}\varphi^{(u-i)}\varphi^{(u-j)} + O(1),$$

$$1\leqslant i,j\leqslant p \tag{4.41}$$

注意，(4.29)式中的 $\dfrac{\partial\gamma_t}{\partial\theta_j}$ 和(4.28)式中的 $\dfrac{\partial\gamma_t}{\partial\varphi_j}$ 除差一负号外完全相似，由此并参照(4.26)和(4.41)式，又易知有

$$E_\beta\left(\frac{\partial l}{\partial\varphi_i}\frac{\partial l}{\partial\theta_j}\right) = -n\sum_{u=\max(i,j)}^{\infty}\varphi^{(u-i)}\theta^{(u-j)} + O(1)$$

$$1\leqslant i\leqslant p,\ 1\leqslant j\leqslant q \tag{4.42}$$

$$E_\beta\left(\frac{\partial l}{\partial \theta_i}\frac{\partial l}{\partial \theta_j}\right) = n\sum_{u=\max(i,j)}^{\infty}\theta^{(u-i)}\theta^{(u-j)} + O(1)$$

$$1 \leqslant i,j \leqslant q \tag{4.43}$$

于是最后可得 $J_n(\beta)$ 的渐近表达式:

$$J_n(\beta) = E_\beta\left(\frac{\partial l}{\partial \beta}\frac{\partial l}{\partial \beta^\tau}\right) = n\begin{pmatrix} J_{\varphi\varphi} & -J_{\varphi\theta} \\ -J_{\varphi\theta}^\tau & J_{\theta\theta}\end{pmatrix} + O(1) \tag{4.44}$$

其中 $J_{\varphi\varphi}$, $J_{\varphi\theta}$, $J_{\theta\theta}$ 分别为 $p \times p$, $p \times q$, $q \times q$ 矩阵，它们的 i 行 j 列元分别为

$$\sum_{u=\max(i,j)}\varphi^{(u-i)}\varphi^{(u-j)},\quad \sum_{u=\max(i,j)}^{\infty}\varphi^{(u-i)}\theta^{(u-j)},\quad \sum_{u=\max(i,j)}\theta^{(u-i)}\theta^{(u-j)}$$

而 $\varphi^{(u)}$ 和 $\theta^{(u)}$ 是由 (4.27) 式定义，因而被参数向量 $\beta^\tau = (\varphi^\tau, \theta^\tau)$ 所唯一确定.

为了求出 $J_n^{-1}(\beta)$ 的渐近表达式(这在实用上是更重要的)，除了 (4.44) 式外，还需要以下两个定理.

定理 2. 设平稳序列 w_t 的对数似然函数 $l(\beta)$ 以 β 为参数向量(其维数为 γ)，其信息阵 $J_n(\beta)$ 存在. 又设在向量 ζ 与 β 之间存在着如下的变换:

$$\zeta^\tau = \zeta^\tau(\beta) = (\zeta_1(\beta), \zeta_2(\beta), \cdots, \zeta_\gamma(\beta))$$

$\zeta(\beta)$ 在 β^0 处的邻域内关于 β 连续可微，而且变换的 Jacobi 阵为可逆阵. 那么在 $\zeta^0 = \zeta(\beta^0)$ 的某个邻域内，此序列以 ζ 为参数的对数似然函数和信息阵可以分别表成

$$l^*(\zeta) = l(\beta(\zeta)) \tag{4.45}$$

$$J_n^*(\zeta) = \left(\frac{\partial \zeta^\tau}{\partial \beta}\right)^{-1} J_n(\beta(\zeta))\left(\frac{\partial \zeta}{\partial \beta^\tau}\right)^{-1} \tag{4.46}$$

其中 $\dfrac{\partial \zeta}{\partial \beta^\tau}$ 是变换的 Jacobi 阵:

$$\frac{\partial \zeta}{\partial \beta^\tau} = \begin{pmatrix} \dfrac{\partial \zeta_1(\beta)}{\partial \beta_1} & \cdots & \dfrac{\partial \zeta_1(\beta)}{\partial \beta_\gamma} \\ & \cdots\cdots & \\ \dfrac{\partial \zeta_\gamma(\beta)}{\partial \beta_1} & \cdots & \dfrac{\partial \zeta_\gamma(\beta)}{\partial \beta_\gamma} \end{pmatrix} \tag{4.47}$$

证明. 在定理的条件下，根据隐函数定理，$\boldsymbol{\zeta}=\boldsymbol{\zeta}(\boldsymbol{\beta})$ 的反函数 $\boldsymbol{\beta}=\boldsymbol{\beta}(\boldsymbol{\zeta})$ 在 $\boldsymbol{\zeta}^0=\boldsymbol{\zeta}(\boldsymbol{\beta}^0)$ 的某个邻域中存在、单值和连续可微，而且

$$\left|\frac{\partial \boldsymbol{\beta}}{\partial \boldsymbol{\zeta}^{\tau}}\right| \neq 0$$

因此，在 $\boldsymbol{\zeta}^0$ 的某个邻域里，对数似然函数可以表成 (4.45) 式，而且

$$\frac{\partial l^*(\boldsymbol{\zeta})}{\partial \boldsymbol{\zeta}}=\frac{\partial \boldsymbol{\beta}^{\tau}}{\partial \boldsymbol{\zeta}} \frac{\partial l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}=\left(\frac{\partial \boldsymbol{\zeta}^{\tau}}{\partial \boldsymbol{\beta}}\right)^{-1} \frac{\partial l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}$$

故 (4.46) 式也在此邻域中成立.

定理 3. 设多项式 $\varphi(B)$ 和 $\theta(B)$ 由一平稳可逆的 ARMA(p, q) 模型所确定，它们彼此没有公共根. 若令

$$\begin{aligned}\lambda(B) &\equiv 1-\lambda_1 B-\lambda_2 B^2-\cdots-\lambda_{p+q} B^{p+q} \\ &=\varphi(B) \theta(B),\end{aligned} \tag{4.48}$$

并记 $\boldsymbol{\lambda}^{\tau}=(\lambda_1, \lambda_2, \cdots, \lambda_{p+q}), \boldsymbol{\beta}^{\tau}=(\boldsymbol{\varphi}^{\tau}, \boldsymbol{\theta}^{\tau})$

那么 Jacobi $\dfrac{\partial \boldsymbol{\lambda}}{\partial \boldsymbol{\beta}^{\tau}}$ 是可逆阵.

证明. 由 λ 的定义可以逐个推出 $\lambda_1, \lambda_2, \cdots, \lambda_{p+q}$ 的具体函数形式如下:

$$\begin{aligned}
&\lambda_1=\varphi_1+\theta_1 \\
&\lambda_2=-\varphi_1 \theta_1+\varphi_2+\theta_2 \\
&\cdots\cdots\cdots\cdots \\
&\lambda_{p+q-2}=-(\varphi_p \theta_{q-2}+\varphi_{p-1} \theta_{q-1}+\varphi_{p-2} \theta_q) \\
&\lambda_{p+q-1}=-(\varphi_p \theta_{q-1}+\varphi_{p-1} \theta_q) \\
&\lambda_{p+q}=-\varphi_p \theta_q
\end{aligned}$$

或统一表为

$$\lambda_k=-\sum_{l=\max(0, k-q)}^{\min(p, k)} \varphi_l \theta_{k-l}=-\sum_{l=\max(0, k-p)}^{\min(q, k)} \varphi_{k-l} \theta_l$$

$$(\varphi_0=\theta_0=-1)$$

易见

$$\frac{\partial\lambda}{\partial\beta^{\tau}}=\left(\begin{array}{cccccc} \multicolumn{3}{c}{\overbrace{\qquad\qquad}^{p}} & \multicolumn{3}{c}{\overbrace{\qquad\qquad}^{q}} \\ 1 & & 0 & 1 & & 0 \\ -\theta_1 & \ddots & & -\varphi_1 & \ddots & \\ \vdots & \ddots & 1 & \vdots & \ddots & 1 \\ -\theta_q & \ddots & -\theta_1 & \vdots & \ddots & -\varphi_1 \\ & \ddots & \vdots & -\varphi_p & & \vdots \\ 0 & & -\theta_q & 0 & \ddots & -\varphi_p \end{array}\right)\begin{array}{l} \\ \left.\begin{array}{c}\\ \\ \\ \end{array}\right\} q \\ \left.\begin{array}{c}\\ \\ \\ \end{array}\right\} p \end{array} \tag{4.49}$$

将这个矩阵的第 i 行 $(1 \leqslant i \leqslant p+q)$ 依次乘以 B^{i-1} 后加到第 $p+q$ 行，使该行变成 $(\theta(B), B\theta(B), \cdots, B^{p-1}\theta(B), \varphi(B), B\varphi(B), \cdots, B^{q-1}\varphi(B))$，其余各行不动．这新矩阵与原矩阵应有相同的秩．若原矩阵不可逆，那么新矩阵也不可逆，即它的 $p+q$ 列是线性相关的，因而存在非全为零的数组 $(\eta_1, \eta_2, \cdots, \eta_p, \xi_1, \xi_2, \cdots, \xi_q)$ 使得

$$\eta_1\theta(B)+\eta_2 B\theta(B)+\cdots+\eta_p B^{p-1}\theta(B)+\xi_1\varphi(B) + \xi_2 B\varphi(B)+\cdots+\xi_q B^{q-1}\varphi(B)=0$$

若记 $\eta(B)=\sum\limits_{j=1}^{p}\eta_j B^{j-1}$，$\xi(B)=-\sum\limits_{j=1}^{q}\xi_j B^{j-1}$，上式即可写作

$$\eta(B)\theta(B)=\xi(B)\varphi(B)$$

因为 $\theta(B)$ 与 $\varphi(B)$ 无公共根，所以存在多项式 $F(B)$ 和 $G(B)$ 使 $\xi(B)=F(B)\theta(B)$，$\eta(B)=G(B)\varphi(B)$．但 $\eta(B)$ 和 $\xi(B)$ 的次数最多分别是 $p-1$ 和 $q-1$ 次．这就导致了矛盾！因此，(4.49)式的矩阵一定是可逆阵．

有了这两个定理，我们就可以对一般的正态 ARMA 序列，求出大样本信息逆阵的表达式．在本节的第二小节，我们已经求出了正态 AR 序列的信息逆阵的渐近形式，下面的目的是要把一般情形化成为这种已经解决的特殊情形．为此，我们考虑以下的正态 AR $(p+q, 0)$ 序列 x_t：

$$\lambda(B)x_t=\varphi(B)\theta(B)x_t=a_t \tag{4.50}$$

并用 ${}^{AR}J_n(\boldsymbol{\lambda})$ 和 ${}^{AR}J_n^*(\boldsymbol{\beta})$ 分别表示 x_t 的以 $\lambda(B)$ 的系数向量 $\boldsymbol{\lambda}$ 为

参数和以 $\boldsymbol{\beta}^{\tau}=(\boldsymbol{\varphi}^{\tau},\boldsymbol{\theta}^{\tau})$ 为参数的信息阵．依定理 3，$\dfrac{\partial\boldsymbol{\lambda}}{\partial\boldsymbol{\beta}^{\tau}}$ 为可逆阵．再由定理 2 的 (4.46) 式，

$$^{\mathrm{AR}}J_n(\boldsymbol{\lambda})=\left(\frac{\partial\boldsymbol{\lambda}^{\tau}}{\partial\boldsymbol{\beta}}\right)^{-1}{}^{\mathrm{AR}}J_n^{*}(\boldsymbol{\beta})\left(\frac{\partial\boldsymbol{\lambda}}{\partial\boldsymbol{\beta}^{\tau}}\right)^{-1}$$

即

$$^{\mathrm{AR}}J_n^{*}(\boldsymbol{\beta})=\left(\frac{\partial\boldsymbol{\lambda}^{\tau}}{\partial\boldsymbol{\beta}}\right){}^{\mathrm{AR}}J_n(\boldsymbol{\lambda})\left(\frac{\partial\boldsymbol{\lambda}}{\partial\boldsymbol{\beta}^{\tau}}\right) \tag{4.51}$$

注意，若用 γ_i^{x} 表 x_t 的自协方差函数，那么仿 (4.28) 和 (4.29) 两式的算法易知

$$\frac{\partial\gamma_i^{x}}{\partial\varphi_j}=\sum_{u=0}^{\infty}\varphi^{(u)}(\gamma_{i+j+u}^{x}+\gamma_{i-j-u}^{x}) \tag{4.52}$$

$$\frac{\partial\gamma_i^{x}}{\partial\theta_j}=\sum_{u=0}^{\infty}\theta^{(u)}(\gamma_{i+j+u}^{x}+\gamma_{i-j-u}^{x}) \tag{4.53}$$

因此与 (4.44) 式相应，有

$$^{\mathrm{AR}}J_n^{*}(\boldsymbol{\beta})=n\begin{pmatrix}J_{\varphi\varphi} & J_{\varphi\theta}\\ J_{\varphi\theta}^{\tau} & J_{\theta\theta}\end{pmatrix}+O(1) \tag{4.54}$$

由此及 (4.44) 式和 (4.51) 式知

$$\begin{aligned}J_n(\boldsymbol{\beta})&=n\begin{pmatrix}J_{\varphi\varphi} & -J_{\varphi\theta}\\ -J_{\varphi\theta}^{\tau} & J_{\theta\theta}\end{pmatrix}+O(1)\\&=\begin{pmatrix}I_p & 0\\ 0 & -I_q\end{pmatrix}{}^{\mathrm{AR}}J_n^{*}(\boldsymbol{\beta})\begin{pmatrix}I_p & 0\\ 0 & -I_q\end{pmatrix}+O(1)\\&=\begin{pmatrix}I_p & 0\\ 0 & -I_q\end{pmatrix}\left(\frac{\partial\boldsymbol{\lambda}^{\tau}}{\partial\boldsymbol{\beta}}\right){}^{\mathrm{AR}}J_n(\boldsymbol{\lambda})\left(\frac{\partial\boldsymbol{\lambda}}{\partial\boldsymbol{\beta}^{\tau}}\right)\begin{pmatrix}I_p & 0\\ 0 & -I_q\end{pmatrix}\\&\quad+O(1)\end{aligned} \tag{4.55}$$

两边取逆，即得

$$\begin{aligned}J_n^{-1}(\beta)=&\begin{pmatrix}I_p & 0\\ 0 & -I_q\end{pmatrix}\left(\frac{\partial\boldsymbol{\lambda}}{\partial\boldsymbol{\beta}^{\tau}}\right)^{-1}{}^{\mathrm{AR}}J_n^{-1}(\boldsymbol{\lambda})\left(\frac{\partial\boldsymbol{\lambda}^{\tau}}{\partial\boldsymbol{\beta}}\right)^{-1}\\&\cdot\begin{pmatrix}I_p & 0\\ 0 & -I_q\end{pmatrix}+o\left(\frac{1}{n}\right)\end{aligned}$$

$$
-\frac{1}{n}\begin{pmatrix} I_p & 0 \\ 0 & -I_q \end{pmatrix}\left(\frac{\partial \boldsymbol{\lambda}}{\partial \boldsymbol{\beta}^\tau}\right)^{-1} M_{p+q}^{*}\left(\frac{\partial \boldsymbol{\lambda}^\tau}{\partial \boldsymbol{\beta}}\right)^{-1}\begin{pmatrix} I_p & 0 \\ 0 & -I_q \end{pmatrix}
$$

$$
+\,o\left(\frac{1}{n}\right) \tag{4.56}
$$

其中 $\frac{\partial \boldsymbol{\lambda}}{\partial \boldsymbol{\beta}^\tau}$ 由(4.49)式给出，而 $p+q$ 阶方阵 M_{p+q}^{*} 则由本附录 §3 的(3.15)式求得，即将那里的 p 维系数向量 $\boldsymbol{\varphi}$ 换为 $\lambda(B)=\varphi(B)\theta(B)$ 的 $p+q$ 维系数向量 $\boldsymbol{\lambda}$ 而得出．这样就给出了正态 ARMA 序列 w_t 的大样本信息逆阵的渐近表达式．

例如，考虑正态 ARMA (1,1) 序列 w_t：

$$
w_t-\varphi_1 w_{t-1}=a_t-\theta_1 a_{t-1}
$$

这时 $\lambda(B)=\varphi(B)\,\theta(B)=1-(\varphi_1+\theta_1)B+\varphi_1\theta_1 B^2$，即 $\boldsymbol{\lambda}=(\varphi_1+\theta_1,-\varphi_1\theta_1)^\tau$，由(4.22)和(4.49)式，

$$
M_{1+1}^{*}=\begin{pmatrix} 1-\varphi_1^2\theta_1^2 & -(\varphi_1+\theta_1)(1-\varphi_1\theta_1) \\ -(\varphi_1+\theta_1)(1-\varphi_1\theta_1) & 1-\varphi_1^2\theta_1^2 \end{pmatrix}
$$

$$
\frac{\partial \boldsymbol{\lambda}}{\partial \boldsymbol{\beta}^\tau}=\begin{pmatrix} 1 & 1 \\ -\theta_1 & -\varphi_1 \end{pmatrix},\quad \left(\frac{\partial \boldsymbol{\lambda}}{\partial \boldsymbol{\beta}^\tau}\right)^{-1}=\frac{1}{\varphi_1-\theta_1}\begin{pmatrix} \varphi_1 & 1 \\ -\theta_1 & -1 \end{pmatrix}
$$

于是由(4.56)式

$$
J_n^{-1}(\boldsymbol{\beta})=\frac{1-\varphi_1\theta_1}{n(\varphi_1-\theta_1)^2}\begin{pmatrix} (1-\varphi_1^2)(1-\varphi_1\theta_1) & (1-\varphi_1^2)(1-\theta_1^2) \\ (1-\varphi_1^2)(1-\theta_1^2) & (1-\theta_1^2)(1-\varphi_1\theta_1) \end{pmatrix}
$$

$$
+\,o\left(\frac{1}{n}\right)
$$

最后我们指出，为了给出以模型全部参数 $\boldsymbol{\alpha}^\tau=(\boldsymbol{\beta}^\tau,\sigma_a^2)=(\boldsymbol{\varphi}^\tau,\boldsymbol{\theta}^\tau,\sigma_a^2)$ 为变量的正态 ARMA 序列的信息逆阵，容易验证此时仍有类似于 AR 序列情形的渐近公式，即

$$
E_a\left(\frac{\partial l_n}{\partial \sigma_a^2}\frac{\partial l_n}{\partial \sigma_a^2}\right)=\frac{n}{2\sigma_a^4}+O(1),\quad E_a\left(\frac{\partial l_n}{\partial \sigma_a^2}\frac{\partial l_n}{\partial \beta_i}\right)=O(1)
$$

所以

$$J_n^{-1}(\alpha) = \left[E_\alpha \frac{\partial l_n}{\partial \alpha} \frac{\partial l_n}{\partial \alpha^\tau} \right]^{-1} = \begin{pmatrix} J_n^{-1}(\beta) & 0 \\ 0 & \frac{2\sigma_a^4}{n} \end{pmatrix} + o\left(\frac{1}{n}\right) \tag{4.57}$$

其中 $J_n^{-1}(\beta)$ 由 (4.56) 式给出.

§5. 正态 AR 序列参数估计的渐近性质

虽然本书所介绍的这种 ARMA 序列分析方法已经得到广泛的应用，但是关于模型参数估计的渐近均方误差性质的研究还不是很完善的. 其原因之一，如在本书第一章 §5 所指出的，由于数理统计学中在使用渐近优效性这一概念时含义不清(见[10])，导致在时间序列分析中关于参数估计的优效性定义的混乱. 这一点已在第四章 §4 中作了进一步的说明. 另一个原因是，在计算模型参数估计的误差方差时，存在一定的困难. 实际上，在写本书的时候，我们尚未见过关于 ARMA 序列参数估计的渐近方差方面的研究资料. 这一节我们将对正态 AR 序列参数估计的渐近方差进行细致的分析，然后证明这种情况的矩估计和精估计都具有渐近最小方差性质. 为了使读者对于参数估计的各种渐近性质的含义和证明都有所了解，我们还将就正态 AR 序列参数估计的其它渐近性质，给出详细的叙述和证明. 这样，对于这一特殊情况的参数估计的渐近性质能够得到比较完善的了解. 最后，我们又把这些结果推广到非正态的 AR 序列的情况. 进一步推广到正态或非正态的 ARMA 序列的情况，这还是有待研究的课题.

下面讨论正态 AR 序列参数估计的渐近性质. 在这种情形，参数的各种估计——矩估计、最小二乘估计以及最小平方和估计等，当样本 N 很大时都是十分接近的，易知它们都有相同的渐近性质. 因此，我们这里只讨论矩估计的渐近性质.

AR$(p,0)$ 序列的模型为

$$\varphi(B)w_t = a_t$$

其中 $\varphi(B) = 1 - \varphi_1 B - \varphi_2 B - \cdots - \varphi_p B^p$. 仍用以前的符

号，设总参数向量 $\boldsymbol{\alpha}^\tau=(\boldsymbol{\varphi}^\tau,\sigma_a^2)$，$\boldsymbol{\varphi}^\tau=(\varphi_1,\varphi_2,\cdots,\varphi_p)$．参数的 Yule-Walker 估计(即矩估计)为 $\hat{\boldsymbol{\alpha}}^\tau=(\hat{\boldsymbol{\varphi}}^\tau,\hat{\sigma}_a^2)$（见第三章 §3）：

$$\hat{\boldsymbol{\varphi}}=\hat{\Gamma}_p^{-1}\hat{\boldsymbol{b}}_p,\ \hat{\sigma}_a^2=\hat{\gamma}_0-\sum_{j=1}^{p}\hat{\varphi}_j\hat{\gamma}_j=\hat{\gamma}_0-\hat{\boldsymbol{\varphi}}^\tau\hat{\boldsymbol{b}}_p \tag{5.1}$$

其中

$$\hat{\Gamma}_p=\begin{pmatrix}\hat{\gamma}_0 & \hat{\gamma}_1 & \cdots & \hat{\gamma}_{p-1}\\ \hat{\gamma}_1 & \hat{\gamma}_0 & \cdots & \hat{\gamma}_{p-2}\\ & \cdots\cdots & & \\ \hat{\gamma}_{p-1} & \hat{\gamma}_{p-2} & \cdots & \hat{\gamma}_0\end{pmatrix},\quad \hat{\boldsymbol{b}}_p=\begin{pmatrix}\hat{\gamma}_1\\ \hat{\gamma}_2\\ \vdots\\ \hat{\gamma}_p\end{pmatrix}$$

$$\hat{\gamma}_k=\frac{1}{N}\sum_{i=1}^{N-k}w_iw_{i+k}$$

定理 1. 正态 AR$(p,0)$ 序列参数的 Yule-Walker 估计 $\hat{\boldsymbol{\alpha}}$ 当样本长度 $N\to\infty$ 时具有以下性质：

〈i〉 $\hat{\boldsymbol{\alpha}}$ 是 $\boldsymbol{\alpha}$ 的相容估计；

〈ii〉 $\hat{\boldsymbol{\alpha}}$ 是 $\boldsymbol{\alpha}$ 的渐近无偏估计；

〈iii〉 $\sqrt{N}\,\tilde{\alpha}=\sqrt{N}(\hat{\alpha}-\alpha)$的分布收敛于正态分布 $\mathcal{N}(0,J^{-1})$ 其中

$$J^{-1}=\begin{pmatrix}M_p & 0\\ 0 & 2\sigma_a^4\end{pmatrix},\ M_p=\sigma_a^2\Gamma_p^{-1} \tag{5.2}$$

〈iv〉 $\hat{\boldsymbol{\alpha}}$ 是 $\boldsymbol{\alpha}$ 的渐近最小方差估计，即

$$E\tilde{\boldsymbol{\alpha}}\tilde{\boldsymbol{\alpha}}^\tau=\frac{1}{N}J^{-1}+o\left(\frac{1}{N}\right)=J_N^{-1}+o\left(\frac{1}{N}\right) \tag{5.3}$$

其中 $J_N=E_\alpha\left(\frac{\partial l_N}{\partial\boldsymbol{\alpha}}\frac{\partial l_N}{\partial\boldsymbol{\alpha}^\tau}\right)=-E\frac{\partial^2 l_N}{\partial\boldsymbol{\alpha}\partial\boldsymbol{\alpha}^\tau}$ 是序列 w_t 的信息阵，l_N 是 w_t 的对数似然函数(注．见前两节)；

〈v〉 $\tilde{\alpha}$ 的所有三阶以上的矩都是 $o\left(\frac{1}{N}\right)$ 无穷小量，即当 $r\geqslant 3$ 时，

$$|E(\tilde{\alpha}_{i_1}\tilde{\alpha}_{i_2}\cdots\tilde{\alpha}_{i_r})|=o\left(\frac{1}{N}\right) \tag{5.4}$$

本节以后的全部内容在于证明这个定理．为此，我们把它分

成几个较小的定理或引理，完成这些定理或引理的证明后，就相当于证明了整个定理.

引理 1. 设

$$\boldsymbol{\phi}_k = \Gamma_k^{-1}\boldsymbol{b}_k, \boldsymbol{b}_k = \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_k \end{pmatrix}, \quad \Gamma_k = \begin{pmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{k-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{k-2} \\ & & \cdots & \\ \gamma_{k-1} & \gamma_{k-2} & \cdots & \gamma_0 \end{pmatrix} \tag{5.5}$$

而且 Γ_{p+1} 是正定阵，那么当 $k \leqslant p$ 时，$\boldsymbol{\phi}_k$ 的诸分量 φ_{kj} $(1 \leqslant j \leqslant k)$ 必满足

$$|\varphi_{kj}| \leqslant 2^{k-1}, \quad 1 \leqslant j \leqslant k \tag{5.6}$$

证明. 由于 Γ_{p+1} 正定，所以存在 $p+1$ 个均值为零的正态随机变量 $y_0, y_1, \cdots, y_p$，其联合分布的协方差阵为 Γ_{p+1}. 由于 y_i 为正态变量，若令

$\tilde{y}_k = \hat{y}_k - y_k = E(y_k | y_1, y_2, \cdots, y_{k-1}) - y_k$，$k > 1$ 则 $y_1, \tilde{y}_2, \tilde{y}_3, \cdots, \tilde{y}_p$ 为 p 个相互独立的随机变量. 记

$\boldsymbol{y}_k^\tau = (y_1, y_2, \cdots, y_k)$，$\boldsymbol{\bar{y}}_k^\tau = (y_k, y_{k-1}, \cdots, y_1)$ 那么(参看第一章 §5)

$$E(y_0|\boldsymbol{y}_p) = E(y_0\boldsymbol{y}_p^\tau)(E\boldsymbol{y}_p\boldsymbol{y}_p^\tau)^{-1}\boldsymbol{y}_p = \boldsymbol{b}_p^\tau\Gamma_p^{-1}\boldsymbol{y}_p = \boldsymbol{\phi}_p^\tau\boldsymbol{y}_p$$

$$= \varphi_{pp}(\hat{y}_p - \tilde{y}_p) + \sum_{j=1}^{p-1} \varphi_{pj}y_j = z - \varphi_{pp}\tilde{y}_p$$

其中 $z = \varphi_{pp}\hat{y}_p + \sum_{j=1}^{p-1} \varphi_{pj}y_j$ 仅是 $y_1, y_2, \cdots, y_{p-1}$ 的线性组合，故与 $\tilde{y}_p$ 独立，所以 $\varphi_{pp} = -E(y_0\tilde{y}_p)(E\tilde{y}_p^2)^{-1}$. 又令

$$\hat{y}_0 = E(y_0|y_1, y_2, \cdots, y_{p-1}) = \boldsymbol{\phi}_{p-1}^\tau\boldsymbol{y}_{p-1}, \tilde{y}_0 = \hat{y}_0 - y_0$$

$\hat{y}_0$ 也是 $y_1, y_2, \cdots, y_{p-1}$ 的线性组合，故也与 $\tilde{y}_p$ 独立. 所以

$$Ey_0\tilde{y}_p = E(\hat{y}_0 - \tilde{y}_0)\tilde{y}_p = -E\tilde{y}_0\tilde{y}_p$$

此外还有

$$E\hat{y}_0^2 = E\boldsymbol{\phi}_{p-1}^\tau\boldsymbol{y}_{p-1}\boldsymbol{y}_{p-1}^\tau\boldsymbol{\phi}_{p-1} = \boldsymbol{\phi}_{p-1}^\tau\Gamma_{p-1}\boldsymbol{\phi}_{p-1} = \boldsymbol{\phi}_{p-1}^\tau\boldsymbol{b}_{p-1}$$

$$E\hat{y}_p^2 = E[E(y_p|y_1, y_2, \cdots, y_{p-1})]^2 = E[E(y_p|y_{p-1}, y_{p-2}, \cdots, y_1)]^2$$

$$= E[E(y_p|\boldsymbol{\bar{y}}_{p-1})]^2 = \boldsymbol{b}_{p-1}^\tau\Gamma_{p-1}^{-1}\boldsymbol{b}_{p-1} = \boldsymbol{\phi}_{p-1}^\tau\boldsymbol{b}_{p-1} = E\hat{y}_0^2$$

从而 $E\tilde{y}_0^2 = Ey_0^2 - E\hat{y}_0^2 = \gamma_0 - E\hat{y}_p^2 = Ey_p^2 - E\hat{y}_p^2 = E\tilde{y}_p^2$. 于是

$$|\varphi_{pp}| = |E(\tilde{y}_0\tilde{y}_p)(E\tilde{y}_0^2)^{-1}| \leqslant (E\tilde{y}_0^2 E\tilde{y}_p^2)^{\frac{1}{2}}(E\tilde{y}_p^2)^{-1} = 1$$

注意，上述推理对于一切 $k \leqslant p$ 都适用，所以

$$|\varphi_{kk}| \leqslant 1, \qquad 1 \leqslant k \leqslant p \tag{5.7}$$

由 $\boldsymbol{\phi}_k$ 的定义(5.5)式和第二章的(2.2.23)式可知

$$\varphi_{kj} = \varphi_{k-1,j} - \varphi_{kk}\varphi_{k-1,k-j},\ j = 1,2,\cdots,k-1$$

于是由(5.7)式和上式，

$$|\varphi_{21}| = |\varphi_{11} - \varphi_{22}\varphi_{11}| \leqslant |\varphi_{11}| + |\varphi_{11}||\varphi_{22}| \leqslant 1 + 1 = 2$$

假定对于 $k \leqslant n$ (5.6)式已成立，那么

$$|\varphi_{n+1,j}| = |\varphi_{nj} - \varphi_{n+1,n+1}\varphi_{n,n+1-j}| \leqslant |\varphi_{nj}| + |\varphi_{n,n+1-j}|$$
$$\leqslant 2^{n-1} + 2^{n-1} = 2^n$$

依归纳法即证得了(5.6)式.

引理 2. 设 w_t 为正态 AR$(p,0)$ 序列，那么，Yule-Walker 估计 $\hat{\boldsymbol{\phi}}$ 的各分量几乎必然有界：

$$\max_{1\leqslant j\leqslant p} |\hat{\varphi}_j| \leqslant 2^{p-1} \tag{5.8}$$

证明：依 $\hat{\gamma}_k$ 和 Γ_k 的定义，可以将 $\hat{\Gamma}_{p+1}$ 写作

$$\hat{\Gamma}_{p+1} = \frac{1}{N}F_N F_N^{\tau}, F_N = \begin{pmatrix} w_1 & w_2 & \cdots & w_N & & & \\ & w_1 & w_2 & \cdots & w_N & & 0 \\ & & \ddots & \ddots & & \ddots & \\ & 0 & & \ddots & \ddots & & \ddots \\ & & & w_1 & w_2 & \cdots & w_N \end{pmatrix}_{(p+1)\times(N+p)}$$

由此可见 $\hat{\Gamma}_{p+1}$ 是非负定的，而且只要 $\boldsymbol{w}_N$ 不是零向量，则 $\hat{\Gamma}_{p+1}$ 就一定为正定. 显然 $\boldsymbol{w}_N = 0$ 的概率为零，而且即使 N 从一变到无穷，所有使 $\boldsymbol{w}_N = 0$ 的概率也只能为零. 因此 $\hat{\Gamma}_{p+1}$ 几乎必然是正定阵. 然后，根据前面的引理 1，(5.8)式几乎必然成立. 这就证明了引理 2.

定理 2. 在定理 1 的条件下，以下结论成立：

(i) 对任意整数 $i_s(1 \leqslant s \leqslant r)$ 和 j_τ $(1 \leqslant \tau \leqslant l)$，只要 $1 \leqslant i_s \leqslant p$，而且 $r + l \geqslant 3$，就有

$$E|\tilde{\varphi}_{i_1}\tilde{\varphi}_{i_2}\cdots\tilde{\varphi}_{i_r}\tilde{\gamma}_{j_1}\tilde{\gamma}_{j_2}\cdots\tilde{\gamma}_{j_l}| = o\left(\frac{1}{N}\right) \tag{5.9}$$

〈ii〉 $\hat{\boldsymbol{\alpha}}$ 是 $\boldsymbol{\alpha}$ 的渐近无偏估计和相容估计，且进一步有

$$\lim_{N\to\infty} E|\tilde{\boldsymbol{\alpha}}| = 0 \tag{5.10}$$

〈iii〉 当 $m \geqslant 3$ 时

$$E|\tilde{\alpha}_{k_1}\tilde{\alpha}_{k_2}\cdots\tilde{\alpha}_{k_m}| = o\left(\frac{1}{N}\right) \tag{5.11}$$

证明. 在(5.1)式中出现的 $\hat{\Gamma}_p$, $\hat{\boldsymbol{b}}_p$, 其被估计值用 Γ_p, $\boldsymbol{b}_p$ 表示，其误差用 $\tilde{\Gamma}_p$,$\tilde{\boldsymbol{b}}_p$ 表示，其它类似的符号不言自明. 为了简化符号，还暂时将下标 p 也略去，即 $\Gamma_p = \Gamma$, $\boldsymbol{b}_p = \boldsymbol{b}$ 等等. 于是

$$\begin{aligned}\tilde{\boldsymbol{\varphi}} = \hat{\boldsymbol{\varphi}} - \boldsymbol{\varphi} &= \hat{\Gamma}^{-1}\hat{\boldsymbol{b}} - \Gamma^{-1}\boldsymbol{b} = \hat{\Gamma}^{-1}\hat{\boldsymbol{b}} - \Gamma^{-1}\hat{\boldsymbol{b}} + \Gamma^{-1}\hat{\boldsymbol{b}} - \Gamma^{-1}\boldsymbol{b} \\ &= \Gamma^{-1}\tilde{\boldsymbol{b}} + (\hat{\Gamma}^{-1} - \Gamma^{-1})\hat{\boldsymbol{b}} = \Gamma^{-1}\tilde{\boldsymbol{b}} - \Gamma^{-1}\tilde{\Gamma}\hat{\Gamma}^{-1}\hat{\boldsymbol{b}} \\ &= \Gamma^{-1}(\tilde{\boldsymbol{b}} - \tilde{\Gamma}\hat{\boldsymbol{\varphi}})\end{aligned} \tag{5.12}$$

现在定义一个 $p\times(p+1)$ 矩阵 $H = (h_{jk})_{(1\leqslant j\leqslant p, 0\leqslant k\leqslant p)}$ 如下：

$$h_{j0} = -\varphi_j, h_{jk} = -\varphi_{j-k} - \varphi_{j+k}, 1\leqslant k\leqslant p \tag{5.13}$$

其中约定

$$\varphi_0 = -1, \varphi_{-i} = \varphi_{p+i} = 0, i > 0$$

相应地可定义 $\hat{H}$（仍约定 $\hat{\varphi}_0 = -1, \hat{\varphi}_{-i} = \hat{\varphi}_{p+i} = 0, i > 0$），那么易知有

$$\tilde{\boldsymbol{b}} - \tilde{\Gamma}\boldsymbol{\varphi} = H\tilde{\boldsymbol{b}}_0,\ \tilde{\boldsymbol{b}} - \tilde{\Gamma}\hat{\boldsymbol{\varphi}} = \hat{H}\tilde{\boldsymbol{b}}_0 \tag{5.14}$$

其中 $\tilde{\boldsymbol{b}}_0 = (\tilde{\gamma}_0, \tilde{\gamma}_1, \cdots, \tilde{\gamma}_p)$. 于是由(5.12)式，可表

$$\tilde{\boldsymbol{\varphi}} = \Gamma^{-1}\hat{H}\tilde{\boldsymbol{b}}_0 \tag{5.15}$$

注意由引理2的(5.8)式和上面的约定，有 $\max\limits_{j,k}|\hat{h}_{jk}| \leqslant \max\limits_{j,k}(|\hat{\varphi}_{j-k}| + |\hat{\varphi}_{j+k}|) \leqslant 2^{p-1} + 2^{p-1} = 2^p$，并依前记 $\Gamma^{-1} = \sigma_a^{-2}M_p = \sigma_a^{-2}(\mu_{ij})$，因此对于 $1\leqslant i\leqslant p$，

$$\begin{aligned}|\tilde{\varphi}_i| &= |\sum_{j=1}^{p}\sum_{k=0}^{p}\sigma_a^{-2}\mu_{ij}\hat{h}_{jk}\tilde{\gamma}_k| \leqslant p\sigma_a^{-2}\max_{i,j}|\mu_{ij}|\max_{j,k}|\hat{h}_{jk}| \\ &\cdot\sum_{k=0}^{p}|\tilde{\gamma}_k| \leqslant k_1\sum_{k=0}^{p}|\tilde{\gamma}_k|\end{aligned} \tag{5.16}$$

再由$\hat{\sigma}_a^2$的定义（见(5.1)式）和$\sigma_a^2=\gamma_0-\boldsymbol{\varphi}^\tau\boldsymbol{b}$（见第二章(7.13)式），

$$\tilde{\sigma}_a^2=\hat{\sigma}_a^2-\sigma_a^2=\hat{\gamma}_0-\hat{\boldsymbol{\varphi}}^\tau\hat{\boldsymbol{b}}-\gamma_0+\boldsymbol{\varphi}^\tau\boldsymbol{b}=\tilde{\gamma}_0-\hat{\boldsymbol{\varphi}}^\tau(\tilde{\boldsymbol{b}}+\boldsymbol{b})+\boldsymbol{\varphi}^\tau\boldsymbol{b}$$
$$=\tilde{\gamma}_0-\hat{\boldsymbol{\varphi}}^\tau\tilde{\boldsymbol{b}}-\boldsymbol{b}^\tau\tilde{\boldsymbol{\varphi}} \tag{5.17}$$

由此及(5.8)，(5.16)式可知

$$|\tilde{\sigma}_a^2|\leqslant|\tilde{\gamma}_0|+2^{p-1}\sum_{j=1}^{p}|\tilde{\gamma}_j|+\sum_{j=1}^{p}|\gamma_j||\tilde{\varphi}_j|\leqslant k_2\sum_{k=0}^{p}|\tilde{\gamma}_k|$$

最后利用附录§2的定理1和2，以及上式和(5.16)式，立刻得到本定理的〈i〉，〈ii〉和〈iii〉的全部证明．其中在证明〈ii〉时还要注意用一下不等式

$$E|\tilde{\alpha}_j|\leqslant\sqrt{E\tilde{\alpha}_j^2}$$

定理3. 在定理1的条件下，Yule-Walker 估计$\hat{\boldsymbol{\alpha}}$的误差方差阵满足不等式

$$E\tilde{\boldsymbol{\alpha}}\tilde{\boldsymbol{\alpha}}^\tau\geqslant J_N^{-1}(\boldsymbol{\alpha})=\frac{1}{N}\begin{pmatrix}M_p & 0\\ 0 & 2\sigma_a^2\end{pmatrix}+o\left(\frac{1}{N}\right)$$

证明．这一定理实际上是附录§4的一、二两小节的推论．根据§4的定理1及其注4可知上式中前一不等式成立，再根据(4.20)式便知上式中后一等式成立．

定理4. 在定理1的条件下，Yule-Walker 估计$\hat{\boldsymbol{\alpha}}$的误差方差阵满足渐近等式

$$E\hat{\boldsymbol{\alpha}}\tilde{\boldsymbol{\alpha}}^\tau=J_N^{-1}+o\left(\frac{1}{N}\right)=\frac{1}{N}\begin{pmatrix}M_p & 0\\ 0 & 2\sigma_a^2\end{pmatrix}+o\left(\frac{1}{N}\right) \tag{5.18}$$

证明．分为以下三步：

〈i〉 求$E(\tilde{\boldsymbol{\varphi}}\tilde{\boldsymbol{\varphi}}^\tau)$．

由(5.12)和(5.9)式，

$$E(\tilde{\boldsymbol{\varphi}}\tilde{\boldsymbol{\varphi}}^\tau)=\Gamma^{-1}E(\tilde{\boldsymbol{b}}-\tilde{\Gamma}(\boldsymbol{\varphi}+\tilde{\boldsymbol{\varphi}}))(\tilde{\boldsymbol{b}}-\tilde{\Gamma}(\boldsymbol{\varphi}+\tilde{\boldsymbol{\varphi}}))^\tau\Gamma^{-1}$$
$$=\Gamma^{-1}E(\tilde{\boldsymbol{b}}-\tilde{\Gamma}\boldsymbol{\varphi})(\tilde{\boldsymbol{b}}-\tilde{\Gamma}\boldsymbol{\varphi})^\tau\Gamma^{-1}+o\left(\frac{1}{N}\right) \tag{5.19}$$

用α_{ij}表示p阶对称阵$E(\tilde{\boldsymbol{b}}-\tilde{\Gamma}\boldsymbol{\varphi})(\tilde{\boldsymbol{b}}-\tilde{\Gamma}\boldsymbol{\varphi})^\tau$的$i$行$j$列元，依附录§2的定理1，对$i,j=1,\cdots,p$（注意$\varphi_0=-1$），

$$a_{ij} = E\left(\sum_{k=0}^{p} \varphi_k \tilde{\gamma}_{i-k}\right)\left(\sum_{l=0}^{p} \varphi_l \tilde{\gamma}_{j-l}\right) = \sum_{k,l=0}^{p} \varphi_k \varphi_l E \tilde{\gamma}_{i-k} \tilde{\gamma}_{j-l}$$

$$= \frac{1}{N} \sum_{k,l=0}^{p} \varphi_k \varphi_l \sum_{t=-\infty}^{\infty} (\gamma_{t+i-k} \gamma_{t+j-l}$$

$$+ \gamma_{t+i-k} \gamma_{t-j+l}) + o\left(\frac{1}{N}\right)$$

$$= \frac{1}{N} \sum_{k,l=0}^{p} \varphi_k \varphi_l \left(\sum_{t'=-\infty}^{\infty} \gamma_{t'+i-j} \gamma_{t'+k-l}\right.$$

$$\left. + \sum_{t=-\infty}^{\infty} \gamma_{t+i-k} \gamma_{t-j+l}\right) + o\left(\frac{1}{N}\right)$$

$$= \frac{1}{N} \sum_{t=-\infty}^{\infty} \sum_{k,l=0}^{p} \varphi_k \varphi_l (\gamma_{t+i-j} \gamma_{t+k-l}$$

$$+ \gamma_{t+i-k} \gamma_{t-j+l}) + o\left(\frac{1}{N}\right)$$

注意因 w_t 是 AR 序列，所以

$$-\sum_{k=0}^{p} \varphi_k \gamma_{\tau+k} = E w_\tau (w_0 - \varphi_1 w_{-1} - \cdots - \varphi_p w_{-p}) = E(w_\tau a_0)$$

$$-\sum_{k=0}^{p} \varphi_k \gamma_{\tau-k} = E w_0 (w_\tau - \varphi_1 w_{\tau-1} - \cdots - \varphi_p w_{\tau-p}) = E(w_0 a_\tau)$$

且当 $t > s$ 时，$E w_s a_t = 0$。于是

$$a_{ij} = -\frac{1}{N} \sum_{t=-\infty}^{\infty} \left[\gamma_{t+i-j} \sum_{l=0}^{p} \varphi_l E\ (w_{t-l} a_0)\right.$$

$$\left. - E(w_{t-j} a_0) E(w_0 a_{t+i})\right] + o\left(\frac{1}{N}\right)$$

$$= \frac{1}{N} \sum_{t=-\infty}^{\infty} \gamma_{t+i-j} E(a_t a_0) + o\left(\frac{1}{N}\right)$$

$$= \frac{\sigma_a^2}{N} \gamma_{i-j} + o\left(\frac{1}{N}\right), i, j = 1, \cdots, p \qquad (5.20)$$

以此代入 (5.19) 式便得

$$E(\tilde{\varphi}\tilde{\varphi}^{\tau})=\Gamma^{-1}\left[\frac{\sigma_a^2}{N}\Gamma+o\left(\frac{1}{N}\right)\right]\Gamma^{-1}+o\left(\frac{1}{N}\right)$$

$$=\frac{\sigma_a^2}{N}\Gamma^{-1}+o\left(\frac{1}{N}\right)=\frac{1}{N}M_p+o\left(\frac{1}{N}\right)\quad(5.21)$$

〈ii〉 求 $E(\varphi\tilde{\sigma}_a^2)$,

由(5.12),(5.17),(5.9)和(5.21)式,

$$E(\varphi\tilde{\sigma}_a^2)=\Gamma^{-1}E(\tilde{b}-\tilde{\Gamma}(\varphi+\tilde{\varphi}))(\tilde{\gamma}_0-(\varphi+\tilde{\varphi})^{\tau}\tilde{b})^{\tau}-E(\tilde{\varphi}\tilde{\varphi}^{\tau})\,b$$

$$=\Gamma^{-1}E(\tilde{b}-\tilde{\Gamma}\varphi)(\tilde{\gamma}_0-\tilde{b}^{\tau}\varphi)-\frac{\sigma_a^2}{N}\Gamma^{-1}b+o\left(\frac{1}{N}\right)$$

$$=\Gamma^{-1}E(\tilde{b}-\tilde{\Gamma}\varphi)(\tilde{\gamma}_0-\tilde{b}^{\tau}\varphi)-\frac{\sigma_a^2}{N}\varphi+o\left(\frac{1}{N}\right)\quad(5.22)$$

若用 η_j 表 p 维向量 $E(\tilde{b}-\tilde{\Gamma}\varphi)(\tilde{\gamma}_0-\tilde{b}^{\tau}\varphi)$ 的第 i 分量,仿〈i〉,由附录 §2 的定理 1,对 $j=1,\cdots,p$,可算出

$$\eta_j=E\left(\sum_{k=0}^{p}\varphi_k\tilde{\gamma}_{j-k}\right)\left(\sum_{l=0}^{p}\varphi_l\tilde{\gamma}_l\right)-\sum_{k,l=0}^{p}\varphi_k\varphi_l E(\tilde{\gamma}_{j-k}\tilde{\gamma}_l)$$

$$=\frac{1}{N}\sum_{k,l=0}^{p}\varphi_k\varphi_l\sum_{t=-\infty}^{\infty}(\gamma_{t+j-k}\gamma_{t+l}+\gamma_{t+j-k}\gamma_{t-l})+o\left(\frac{1}{N}\right)$$

$$=\frac{1}{N}\sum_{t=-\infty}^{\infty}\sum_{k,l=0}^{p}\varphi_k\varphi_l(\gamma_{t+j-k}\gamma_{t+l}+\gamma_{t+j}\gamma_{t+k-l})+o\left(\frac{1}{N}\right)$$

$$=\frac{1}{N}\sum_{t=-\infty}^{\infty}[E(w_0a_{t+j})E(w_ta_0)+\gamma_{t+j}E(a_ta_0)]+o\left(\frac{1}{N}\right)$$

$$=\frac{\sigma_a^2}{N}\gamma_j+o\left(\frac{1}{N}\right)\quad(5.23)$$

代入(5.22)式即得

$$E(\tilde{\varphi}\tilde{\sigma}_a^2)=\frac{\sigma_a^2}{N}\Gamma^{-1}b-\frac{\sigma_a^2}{N}\Gamma^{-1}b+o\left(\frac{1}{N}\right)=o\left(\frac{1}{N}\right)\quad(5.24)$$

〈iii〉 求 $E(\tilde{\sigma}_a^2)^2$.

由(5.17),(5.24),(5.9),(5.12),(5.23)等式,

$$E(\tilde{\sigma}_a^2)^2=E\tilde{\sigma}_a^2(\tilde{\gamma}_0-(\varphi+\tilde{\varphi})^{\tau}\tilde{b})^{\tau}-E\,\tilde{\sigma}_a^2\tilde{\varphi}^{\tau}b$$

$$= E(\tilde{\gamma}_0 - \boldsymbol{\varphi}^\tau \tilde{\boldsymbol{b}} - \boldsymbol{b}^\tau \tilde{\boldsymbol{\varphi}})(\tilde{\gamma}_0 - \boldsymbol{\varphi}^\tau \tilde{\boldsymbol{b}})^\tau + o\left(\frac{1}{N}\right)$$

$$= E(\tilde{\gamma}_0 - \tilde{\boldsymbol{b}}^\tau \boldsymbol{\varphi})^2 - \boldsymbol{b}^\tau \Gamma^{-1} E(\tilde{\boldsymbol{b}} - \tilde{\Gamma}\boldsymbol{\varphi})(\tilde{\gamma}_0 - \tilde{\boldsymbol{b}}^\tau \boldsymbol{\varphi}) + o\left(\frac{1}{N}\right)$$

$$= E(\tilde{\gamma}_0 - \tilde{\boldsymbol{b}}^\tau \boldsymbol{\varphi})^2 - \frac{\sigma_a^2}{N} \boldsymbol{\varphi}^\tau \boldsymbol{b} + o\left(\frac{1}{N}\right) \tag{5.25}$$

又仿上易知

$$E(\tilde{\gamma}_0 - \tilde{\boldsymbol{b}}^\tau \boldsymbol{\varphi})^2 = E\left(\sum_{k=0}^{p} \varphi_k \tilde{\gamma}_k\right)^2 = \sum_{k,l=0}^{p} \varphi_k \varphi_l E\tilde{\gamma}_k \tilde{\gamma}_l$$

$$= \frac{1}{N} \sum_{k,l=0}^{p} \varphi_k \varphi_l \sum_{t=-\infty}^{\infty} (\gamma_{t+k}\gamma_{t+l} + \gamma_{t+k}\gamma_{t-l}) + o\left(\frac{1}{N}\right)$$

$$= \frac{1}{N} \sum_{t=-\infty}^{\infty} \sum_{k,l=0}^{p} \varphi_k \varphi_l (\gamma_{t+k-l}\gamma_t + \gamma_{t+k}\gamma_{t-l}) + o\left(\frac{1}{N}\right)$$

$$= \frac{1}{N} \sum_{t=-\infty}^{\infty} [\gamma_t E(a_t a_0) + E(w_t a_0) E(w_0 a_t)] + o\left(\frac{1}{N}\right)$$

$$= \frac{1}{N} [\gamma_0 \sigma_a^2 + (E w_0 a_0)^2] + o\left(\frac{1}{N}\right) = \frac{\sigma_a^2}{N} (\gamma_0 + \sigma_a^2)$$

$$+ o\left(\frac{1}{N}\right)$$

将此代入 (5.25) 式，并注意 $\sigma_a^2 = \gamma_0 - \boldsymbol{\varphi}^\tau \boldsymbol{b}$ 即得

$$E(\tilde{\sigma}_a^2)^2 = \frac{\sigma_a^2}{N}(\gamma_0 + \sigma_a^2) - \frac{\sigma_a^2}{N}(\gamma_0 - \sigma_a^2) + o\left(\frac{1}{N}\right)$$

$$= \frac{2\sigma_a^4}{N} + o\left(\frac{1}{N}\right) \tag{5.26}$$

将 (5.21)，(5.24)，(5.26) 式联合起来，便证明了 (5.18) 式.

定理 5. 在定理 1 的条件下，令

$$\tilde{\boldsymbol{\varphi}}_k = \hat{\Gamma}_k^{-1} \hat{\boldsymbol{b}}_k \tag{5.27}$$

其中 $\hat{\boldsymbol{b}}_k^\tau = (\hat{\gamma}_1, \hat{\gamma}_2, \cdots, \hat{\gamma}_k)$，$\hat{\Gamma}_k$ 是 k 阶方阵，其 i 行 j 列元为 $\hat{\gamma}_{i-j}$，那么当 $k \geqslant p$ 时

$$E(\tilde{\boldsymbol{\varphi}}_k \tilde{\boldsymbol{\varphi}}_k^\tau) = E(\tilde{\boldsymbol{\varphi}}_k - \boldsymbol{\phi}_k)(\tilde{\boldsymbol{\varphi}}_k - \boldsymbol{\phi}_k)^\tau = \frac{1}{N} M_k + o\left(\frac{1}{N}\right) \tag{5.28}$$

其中 $\boldsymbol{\phi}_k = \Gamma_k^{-1}\boldsymbol{b}_k$ 由 (5.5) 式定义.

证明. 在定理 4 (5.21) 式的推导过程中,如果我们把 w_t 的模型看成是退化的 AR $(k,0)$ 序列 $(k > p)$, 其模型参数为 $\boldsymbol{\phi}_k^{\tau} = (\varphi_1, \varphi_2, \cdots, \varphi_p, 0, 0, \cdots, 0) = \Gamma_k^{-1}\boldsymbol{b}_k$, 那么整个证明将毫无改变地适用于估计 $\hat{\boldsymbol{\phi}}_k$, 而且仍有类似的结果,即

$$E\tilde{\boldsymbol{\varphi}}_k\tilde{\boldsymbol{\varphi}}_k^{\tau} = \frac{\sigma_a^2}{N}\Gamma_k^{-1} + o\left(\frac{1}{N}\right) = \frac{1}{N}M_k + o\left(\frac{1}{N}\right)$$

关于 AR 序列的 M_k, 其诸元的具体表达式已经在附录 §3 里给出了. 在此,我们特别关心 $\hat{\boldsymbol{\phi}}_k$ 中序数大于 p 的诸分量 $\hat{\varphi}_{kj}(j > p)$ 的误差方差,这时 $\varphi_{kj} = 0$. 依附录 (3.17) 式和 (5.28) 式便知,

$$E\tilde{\varphi}_{kj}^2 = E\hat{\varphi}_{kj}^2 = \frac{1}{N}\sum_{l=0}^{\min(k-j,p)}\varphi_l^2 + o\left(\frac{1}{N}\right), \quad k > p, j > p \qquad (5.29)$$

可见 $E\tilde{\varphi}_{kk}^2 \approx \frac{1}{N}$, $k > p$, 且当 $p < j \leqslant k - p$ 时,

$$E\tilde{\varphi}_{kj}^2 \approx \frac{1}{N}\sum_{l=0}^{p}\varphi_l^2 = \frac{1}{N}(1 + \varphi_1^2 + \cdots + \varphi_p^2) \qquad (5.30)$$

又知当 $p < i, j \leqslant k$ 且 $|i - j| > p$ 时, $E\tilde{\varphi}_{ki}\tilde{\varphi}_{kj} = o\left(\frac{1}{N}\right)$.

在本书第三章中,除用到 $E\tilde{\varphi}_{kk}^2$ 的渐近表示式外, 还用到下面的结果:

定理 6. 在定理 1 的条件下, 并采用定理 5 中的记号,那么当 $n > m \geqslant p$ 时

$$E\tilde{\varphi}_{mm}\tilde{\varphi}_{nn} = E\hat{\varphi}_{mm}\hat{\varphi}_{nn} = o\left(\frac{1}{N}\right) \qquad (5.31)$$

证明. 注意,对于 AR$(p,0)$ 序列, 无论偏相关函数,还是样本偏相关函数, 它们都满足第二章 (2.2.23) 式的递推关系. 只是对于样本偏相关函数, 其中的诸量应以相应的样本值代替. 我们这里主要用到 (2.2.23) 式中的最后一组公式, 即对每个 $k \geqslant 1$,

$$\hat{\varphi}_{k+1,j} = \hat{\varphi}_{kj} - \hat{\varphi}_{k+1,k+1}\hat{\varphi}_{k,k-j+1}, \quad 1 \leqslant j \leqslant k \qquad (5.32)$$

$$\varphi_{k+1,j} = \varphi_{kj} - \varphi_{k+1,k+1}\varphi_{k,k-j+1}, \quad 1 \leqslant j \leqslant k \qquad (5.33)$$

用 (5.32) 式的两边分别减 (5.33) 式两边，当 $k \geqslant p$ 时便得到（注意 $\varphi_{k+1,k+1}=0$，在 $k \geqslant p$ 时）

$$\tilde{\varphi}_{k+1,j}=\tilde{\varphi}_{kj}-(\hat{\varphi}_{k+1,k+1}\hat{\varphi}_{k,k-j+1}-\varphi_{k+1,k+1}\varphi_{k,k-j+1})$$
$$=\tilde{\varphi}_{kj}-\tilde{\varphi}_{k+1,k+1}\hat{\varphi}_{k,k-j+1} \tag{5.34}$$

以 $\tilde{\varphi}_{k+1,k+1}$ 同乘上式两边再求各自均值得

$$E\tilde{\varphi}_{k+1,k+1}\tilde{\varphi}_{k+1,j}=E\tilde{\varphi}_{k+1,k+1}\tilde{\varphi}_{kj}-E\tilde{\varphi}^2_{k+1,k+1}\hat{\varphi}_{k,k-j+1}$$
$$=E\tilde{\varphi}_{k+1,k+1}\tilde{\varphi}_{kj}-\varphi_{k,k-j+1}E\hat{\varphi}^2_{k+1,k+1}-E\tilde{\varphi}^2_{k+1,k+1}\tilde{\varphi}_{k,k-j+1} \tag{5.35}$$

根据定理 2 的结果和证明，不难看出有

$$|E\tilde{\varphi}^2_{k+1,k+1}\tilde{\varphi}_{k,k-j+1}|=o\left(\frac{1}{N}\right)$$

以此代入 (5.35) 式，移项即有

$$E\tilde{\varphi}_{k+1,k+1}\tilde{\varphi}_{kj}=E\tilde{\varphi}_{k+1,k+1}\tilde{\varphi}_{k+1,j}+\varphi_{k,k-j+1}E\hat{\varphi}^2_{k+1,k+1}+o\left(\frac{1}{N}\right)$$

注意

$$\varphi_{k,k-j+1}=\begin{cases}0, & k-j+1>p\\ \varphi_{k-j+1}, & k-j+1\leqslant p\end{cases}$$

应用 (5.29) 式和附录 §3 的 (3.17) 式分别计算出前一式右边的前两项，可得

$$E\tilde{\varphi}_{k+1,k+1}\tilde{\varphi}_{kj}=-\frac{1}{N}\varphi_{k,k-j+1}+\frac{1}{N}\varphi_{k,k-j+1}+o\left(\frac{1}{N}\right)$$
$$=o\left(\frac{1}{N}\right)\quad k\geqslant p,1\leqslant j\leqslant k \tag{5.36}$$

再在 (5.34) 式两边乘以 $\tilde{\varphi}_{k+2,k+2}$，然后求各自均值，又有

$$E\tilde{\varphi}_{k+2,k+2}\tilde{\varphi}_{k+1,j}=E\tilde{\varphi}_{k+1,k+1}\tilde{\varphi}_{kj}-E\tilde{\varphi}_{k+1,k+1}\tilde{\varphi}_{k+2,k+2}\hat{\varphi}_{k,k-j+1}$$
$$=E\tilde{\varphi}_{k+2,k+2}\tilde{\varphi}_{kj}-\varphi_{k,k-j+1}E\tilde{\varphi}_{k+1,k+1}\tilde{\varphi}_{k+2,k+2}+o\left(\frac{1}{N}\right)$$

利用 (5.36) 式（将 k 换成 $k+1$），由上式便得

$$E\tilde{\varphi}_{k+2,k+2}\tilde{\varphi}_{kj}=E\tilde{\varphi}_{k+2,k+2}\tilde{\varphi}_{k+1,j}+\varphi_{k,k-j+1}E\tilde{\varphi}_{k+1,k+1}\tilde{\varphi}_{k+2,k+2}$$
$$+o\left(\frac{1}{N}\right)=o\left(\frac{1}{N}\right),\quad k\geqslant p,\quad 1\leqslant j\leqslant k$$

依归纳法不难得到，对任意正整数 τ，都有

$$E\tilde{\varphi}_{k+\tau,k+\tau}\tilde{\varphi}_{kj}=o\left(\frac{1}{N}\right),\ k\geqslant p,1\leqslant j\leqslant k$$

特别取 $j=k=m$，$k+\tau=n$ 便得到(5.31)式．定理证毕．

(5.29)和(5.31)式说明，AR$(p,0)$序列的样本偏相关函数 $\hat{\varphi}_{kk}$，在 $k>p$ 以后是渐近无关的．

下面我们来研究AR序列参数的Yule-Walker估计的渐近分布．为此先证明以下的引理．

引理3. 设 r 阶方阵 Q_n 为随机矩阵，当 $n\to\infty$ 时，Q_n 依概率收敛于常数阵 Q，即 Q_n 的每个元 $q_{ij}^{(n)}$ 依概率收敛于常数 q_{ij} $(1\leqslant i,j\leqslant r)$；又设 $\boldsymbol{x}_n$ 为 r 维随机向量序列，当 $n\to\infty$ 时，$\boldsymbol{x}_n$ 的分布渐近于正态 $\mathscr{N}(0,A)$ 分布．那么，随机向量 $\boldsymbol{y}_n=Q_n\boldsymbol{x}_n$ 的分布渐近于正态 $\mathscr{N}(0,QAQ^\tau)$ 分布．

证明．〈i〉 设 $\boldsymbol{\alpha}$ 为任意的 r 维实向量，

$$\boldsymbol{\eta}_n=Q_n^\tau\boldsymbol{\alpha},\boldsymbol{\eta}_n^\tau\boldsymbol{x}_n=\boldsymbol{\alpha}^\tau Q_n\boldsymbol{x}_n=\boldsymbol{\alpha}^\tau\boldsymbol{y}_n$$

于是当 $n\to\infty$ 时，$\boldsymbol{\eta}_n$ 依概率收敛于常向量 $\boldsymbol{\eta}=Q^\tau\boldsymbol{\alpha}$

〈ii〉 $\boldsymbol{x}_n$ 的分布渐近于正态 $\mathscr{N}(0,A)$，等价于 $\boldsymbol{x}_n$ 的特征函数收敛于 $\mathscr{N}(0,A)$ 的特征函数，即当 $n\to\infty$ 时

$$Ee^{i\boldsymbol{\xi}^\tau\boldsymbol{x}_n}\to e^{-\frac{1}{2}\boldsymbol{\xi}^\tau A\boldsymbol{\xi}}$$

对任意 r 维实向量 $\boldsymbol{\xi}$ 都成立．若以 $t\boldsymbol{\xi}$（t 为任意实数）代替 $\boldsymbol{\xi}$，由上式有

$$Ee^{it\boldsymbol{\xi}^\tau\boldsymbol{x}_n}\to e^{-\frac{t^2}{2}\boldsymbol{\xi}^\tau A\boldsymbol{\xi}}\qquad(n\to\infty)$$

它说明随机变量列 $\boldsymbol{\xi}^\tau\boldsymbol{x}_n$ 的特征函数收敛于正态 $\mathscr{N}(0,\boldsymbol{\xi}^\tau A\boldsymbol{\xi})$ 的特征函数，从而 $\boldsymbol{\xi}^\tau\boldsymbol{x}_n$ 的分布渐近于正态 $\mathscr{N}(0,\boldsymbol{\xi}^\tau A\boldsymbol{\xi})$

〈iii〉 令

$$\phi(x)=(2\pi)^{-r/2}|A|^{-\frac{1}{2}}\int_{|\boldsymbol{\xi}|>x}e^{-\frac{1}{2}\boldsymbol{\xi}^\tau A^{-1}\boldsymbol{\xi}}d\boldsymbol{\xi}$$

其中 $\boldsymbol{\xi}^\tau=(\xi_1,\xi_2,\cdots,\xi_r)$ 为 r 维向量，并约定记 $|\boldsymbol{\xi}|=\sum_{i=1}^{r}|\xi_i|$．那么，任给 $\varepsilon>0$，必存在 $K_1>0$，使 $\phi(K_1)<\frac{\varepsilon}{2}$．因为 $\boldsymbol{x}_n$ 的分

布渐近于正态 $\mathscr{N}(0,A)$，所以当 n 大于某一 n_0 后，

$$P[|\boldsymbol{x}_n| \geqslant k_1] < \psi(k_1) + \frac{\varepsilon}{2} < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

我们还可选取 $K_2 > 0$，使

$$P[|\boldsymbol{x}_n| \geqslant K_2] < \varepsilon, 1 \leqslant n \leqslant n_0$$

再取 $K = \max(K_1, K_2)$，则有

$$P[|\boldsymbol{x}_n| \geqslant K] < \varepsilon, 1 \leqslant n < \infty$$

〈iv〉记 $\tilde{\boldsymbol{\eta}}_n = \boldsymbol{\eta}_n - \boldsymbol{\eta}$，易见 $|\tilde{\boldsymbol{\eta}}_n^{\tau}\boldsymbol{x}_n| \leqslant |\tilde{\boldsymbol{\eta}}_n|\,|\boldsymbol{x}_n|$. 于是由〈iii〉，任给 $\delta > 0$，

$$\begin{aligned}P[|\tilde{\boldsymbol{\eta}}_n^{\tau}\boldsymbol{x}_n| \geqslant \delta] &\leqslant P[|\tilde{\boldsymbol{\eta}}_n||\boldsymbol{x}_n| \geqslant \delta]\\ &= P[|\tilde{\boldsymbol{\eta}}_n||\boldsymbol{x}_n| \geqslant \delta, |\boldsymbol{x}_n| < K] + P[|\tilde{\boldsymbol{\eta}}_n||\boldsymbol{x}_n| \geqslant \delta, |\boldsymbol{x}_n|\\ &\geqslant K] \leqslant P[|\tilde{\boldsymbol{\eta}}_n|K \geqslant \delta] + P[|\boldsymbol{x}_n| \geqslant K] \leqslant P\left[|\tilde{\boldsymbol{\eta}}_n| \geqslant \frac{\delta}{K}\right]\\ &\quad + \varepsilon\end{aligned}$$

在〈i〉中已指出 $\boldsymbol{\eta}_n$ 依概率收敛于 $\boldsymbol{\eta}$，因而 $|\tilde{\boldsymbol{\eta}}_n|$ 依概率收敛于 0。于是在上式中先令 $n \to \infty$，再令 $\varepsilon \to 0$，即得

$$\lim_{n\to\infty} P[|\tilde{\boldsymbol{\eta}}_n^{\tau}\boldsymbol{x}_n| \geqslant \delta] = 0$$

这就证明了 $\tilde{\boldsymbol{\eta}}_n^{\tau}\boldsymbol{x}_n$ 也依概率收敛于 0.

〈v〉注意，任给 $\delta > 0$ 和实数 x，

$$\begin{aligned}P[\boldsymbol{\eta}_n^{\tau}\boldsymbol{x}_n < x] &= P[\boldsymbol{\eta}_n^{\tau}\boldsymbol{x}_n < x, \boldsymbol{\eta}^{\tau}\boldsymbol{x}_n < x + \delta]\\ &\quad + P[\boldsymbol{\eta}_n^{\tau}\boldsymbol{x}_n < x, \boldsymbol{\eta}^{\tau}\boldsymbol{x}_n \geqslant x + \delta]\\ &\leqslant P[\boldsymbol{\eta}^{\tau}\boldsymbol{x}_n < x + \delta] + P[|\tilde{\boldsymbol{\eta}}_n^{\tau}\boldsymbol{x}_n| \geqslant \delta]\end{aligned}$$

$$\begin{aligned}P[\boldsymbol{\eta}^{\tau}\boldsymbol{x}_n < x - \delta] &= P[\boldsymbol{\eta}^{\tau}\boldsymbol{x}_n < x - \delta, \boldsymbol{\eta}_n^{\tau}\boldsymbol{x}_n < x]\\ &\quad + P[\boldsymbol{\eta}^{\tau}\boldsymbol{x}_n < x - \delta, \boldsymbol{\eta}_n^{\tau}\boldsymbol{x}_n \geqslant x]\\ &\leqslant P[\boldsymbol{\eta}_n^{\tau}\boldsymbol{x}_n < x] + P[|\tilde{\boldsymbol{\eta}}_n^{\tau}\boldsymbol{x}_n| \geqslant \delta]\end{aligned}$$

联合起来就有

$$\begin{aligned}P[\boldsymbol{\eta}^{\tau}\boldsymbol{x}_n < x - \delta] - P[|\tilde{\boldsymbol{\eta}}_n^{\tau}\boldsymbol{x}_n| \geqslant \delta] &\leqslant P[\boldsymbol{\eta}_n^{\tau}\boldsymbol{x}_n < x]\\ &\leqslant P[\boldsymbol{\eta}^{\tau}\boldsymbol{x}_n < x + \delta] + P[|\tilde{\boldsymbol{\eta}}_n^{\tau}\boldsymbol{x}_n| \geqslant \delta]\end{aligned}$$

由〈ii〉，$\boldsymbol{\eta}^{\tau}\boldsymbol{x}_n$ 的分布渐近于正态 $N(0, \boldsymbol{\eta}^{\tau}A\boldsymbol{\eta})$. 于是由上式和〈iv〉得到

$$\varliminf_{n\to\infty} P[\boldsymbol{\eta}^\tau \boldsymbol{x}_n < x-\delta] \leqslant \varliminf_{n\to\infty} P[\boldsymbol{\eta}_n^\tau \boldsymbol{x}_n < x] \leqslant \varlimsup_{n\to\infty} P[\boldsymbol{\eta}_n^\tau \boldsymbol{x}_n < x] \leqslant \lim_{n\to\infty} P[\boldsymbol{\eta}^\tau \boldsymbol{x}_n < x+\delta]$$

再令 $\delta \to 0$，即知

$$\lim_{n\to\infty} P[\boldsymbol{\eta}_n^\tau \boldsymbol{x}_n < x] = \lim_{n\to\infty} P[\boldsymbol{\eta}^\tau \boldsymbol{x}_n < x]$$

即 $\boldsymbol{\eta}_n^\tau \boldsymbol{x}_n$ 的分布也渐近于正态 $\mathcal{N}(0, \boldsymbol{\eta}^\tau A \boldsymbol{\eta})$

〈vi〉在〈i〉中已知 $\boldsymbol{\eta} = Q^\tau \boldsymbol{u}$，$\boldsymbol{\eta}_n^\tau \boldsymbol{x}_n = \boldsymbol{u}^\tau Q_n \boldsymbol{x}_n = \boldsymbol{u}^\tau \boldsymbol{y}_n$，而 $\boldsymbol{u}$ 是任意的 r 维实向量. 由〈v〉知

$$E e^{i\boldsymbol{u}^\tau Q_n \boldsymbol{x}_n} = E e^{i\boldsymbol{u}^\tau \boldsymbol{y}_n} = E e^{i\boldsymbol{\eta}_n^\tau \boldsymbol{x}_n} \to e^{-\frac{1}{2}\boldsymbol{\eta}^\tau A \boldsymbol{\eta}} = e^{-\frac{1}{2}\boldsymbol{u}^\tau Q A Q^\tau \boldsymbol{u}}$$

而这又等价于 $\boldsymbol{y}_n = Q_n \boldsymbol{x}_n$ 的分布渐近于正态 $\mathcal{N}(0, QAQ^\tau)$. 引理 3 证毕.

定理 7. 在定理 1 的条件下，设 $\boldsymbol{\alpha}$ 的 Yule-Walker 估计为 $\hat{\boldsymbol{\alpha}}$，那么 $\sqrt{N}\,\tilde{\boldsymbol{\alpha}}$ 的分布渐近于正态 $\mathcal{N}(0, J^{-1})$. J^{-1} 的定义见(5.2)式.

证明. 根据前面推导的(5.14)和(5.17)式,

$$\sqrt{N}\,\tilde{\boldsymbol{\varphi}} = \Gamma^{-1}\hat{H}\,\sqrt{N}\,\tilde{\boldsymbol{b}}_0$$

$$\begin{aligned}\sqrt{N}\,\tilde{\sigma}_a^2 &= (\tilde{\gamma}_0 - \hat{\boldsymbol{\varphi}}^\tau \tilde{\boldsymbol{b}} - \boldsymbol{b}^\tau \Gamma^{-1}\hat{H}\tilde{\boldsymbol{b}}_0)\sqrt{N} \\ &= -[(\hat{\varphi}_0, \hat{\boldsymbol{\varphi}}^\tau) + \boldsymbol{\varphi}^\tau \hat{H}]\sqrt{N}\,\tilde{\boldsymbol{b}}_0\end{aligned}$$

其中 $\hat{\varphi}_0 = -1$. 在定理 2 中已经证明了 $E|\tilde{\boldsymbol{\varphi}}| \to 0$，所以 $\hat{\boldsymbol{\varphi}}$ 和 $\hat{H}$ 分别依概率收敛于 $\boldsymbol{\varphi}$ 和 H，其中 H 的元的定义如(5.13)式. 于是上面两式可统一表为

$$\sqrt{N}\,\tilde{\boldsymbol{\alpha}} = \begin{pmatrix} \Gamma^{-1}\hat{H} \\ -(\hat{\varphi}_0, \hat{\boldsymbol{\varphi}}^\tau) - \boldsymbol{\varphi}^\tau \hat{H} \end{pmatrix} \sqrt{N}\,\boldsymbol{b}_0 = Q_N \sqrt{N}\,\tilde{\boldsymbol{b}}_0$$

$$Q_N \xrightarrow[N\to\infty]{P} Q = \begin{pmatrix} \Gamma^{-1}H \\ -(\varphi_0, \boldsymbol{\varphi}^\tau) - \boldsymbol{\varphi}^\tau H \end{pmatrix}$$

其中的收敛记号“$\xrightarrow{P}$”表示依概率收敛. 依附录 §2 的定理 3，$\sqrt{N}\,\tilde{\boldsymbol{b}}_0$ 的分布渐近于正态 $\mathcal{N}(0, G)$，其中 G 的元由(2.22)式给出. 再由引理 3 即知，$\sqrt{N}\,\tilde{\boldsymbol{\alpha}}$ 的分布渐近于正态 $\mathcal{N}(0, QGQ^\tau)$，这里

$$QGQ^\tau = \begin{pmatrix} \Gamma^{-1}H \\ -(\varphi_0, \boldsymbol{\varphi}^\tau) - \boldsymbol{\varphi}^\tau H \end{pmatrix} G\,(H^\tau \Gamma^{-1}, -(\varphi_0, \boldsymbol{\varphi}^\tau)^\tau - H^\tau \boldsymbol{\varphi})$$

$$= \begin{pmatrix} \Gamma^{-1}HGH^{\tau}\Gamma^{-1} - \Gamma^{-1}HG[(\varphi_0, \boldsymbol{\varphi}^{\tau})^{\tau} + H^{\tau}\boldsymbol{\varphi}] \\ -[(\varphi_0, \boldsymbol{\varphi}^{\tau}) + \boldsymbol{\varphi}^{\tau}H]GH^{\tau}\Gamma^{-1}[(\varphi_0, \boldsymbol{\varphi}^{\tau}) + \boldsymbol{\varphi}^{\tau}H]G[(\varphi_0, \boldsymbol{\varphi}^{\tau})^{\tau} + H^{\tau}\boldsymbol{\varphi}] \end{pmatrix}$$

仿定理 4 证明中的方法计算出上述矩阵中的每一元（这里不再赘述了），其结果表明：

$$QGQ^{\tau} = \begin{pmatrix} M_p & 0 \\ 0 & 2\sigma_a^2 \end{pmatrix} = J^{-1}$$

从而证明了$\sqrt{N}\,\tilde{a}$ 的分布渐近于正态 $\mathcal{N}(0, J^{-1})$. 于是定理全部得证.

参考文献

[1] P. Whittle, Estimation and information in stationary time series, Arkiv för Matematik Band 2 nr32, 1952.

[2] J. L. Doob, Stochastic Processes, New York, 1953.

[3] E. Lukacs, Characteristic Functions, London, 1960.

[4] J. Durbin, Estimation of parameters in time series regression models, Journal of the Royal Statistical Society, Series B, Vol. 22. p. 139- 153, 1960.

[5] U. Grenander and M. Rosenblatt, 平稳时间序列的统计分析，郑绍谦等译，上海科学技术出版社，1962.

[6] M. Fisz, 概率论及数理统计，王福保译，上海科学技术出版社，1962.

[7] J. H. Laning and R. H. Battin, 自动控制中的随机过程，徐其例译，科学出版社，1962.

[8] M. Loève, Probability Theory, New York, 1963.

[9] K. J. Åström and T. Bohlin, Numerical identification of linear denamical systems from normal operating records, presented at the IFAC Symp. Theory of Self-Adoptive Systems, Teddington, England, 1965.

[10] H. Cramèr, 统计学数学方法，魏宗舒等译，上海科学技术出版社，1966.

[11] J. Capon, R. J. Geenfield and R. J. Kolker, Multidimensional maximum-Likelihood processing of a large Aperture Seismic Array, *Proc. of the IEEE*, Vol. 55, No. 2, 1967.

[12] G. M. Jenkins. and D. G. Watts, Spectral Analysis. and its applications, London, 1969.

[13] K. Toman, The spectral shifts of truncated sinusoids, *J. Geophysical Resarch*, Vol. 70, No. 7, p 1749—1750.

[14] G. E. P. Box and G. M. Jenkins, Time Series, Forecasting and Control, San Francisco, 1970.

[15] E. J. Hannan, Multiple Time Sevies, New York, 1970.

[16] E. J. Hannan, The asymptotic theory of linear time series models, *J. Appl. Probability*, Vol. 8, p 130—145, 1973.

[17] 张建中，蒙特卡洛方法(I)，数学的实践与认识，1974年1月.

[18] 中国科学院数学研究所统计组，常用数理统计方法，科学出版社，1974.

[19] *IEEE Transaction on Automatic Control*, Vol. AC-19 No. 6, 1974年12月.

[20] 中国科学院数学研究所概率组，离散时间系统滤波的数学方法，国防工业出版社，1975.

[21] 杜金观，潘一民，协方差可分离随机序列的新息与滤波，数学学报，第20卷第1期 1977年3月.

[22] Benjamin. C. Kuo, Analysis and Synthesis of Sampled-Data Control Systems. Englewood Cliffs N. J. Prentice-Hall. 1963.

[23] H. Akaike, On entropy maxmization principle, In applications of Statistics, (P. R. Krishnaiah, ed.) p. 27—41, 1977.

[24] An Hong-Zhi, The estimation of the prediction error variance, *Stochastic Processes and their Applications*, **13**, p. 39—43, 1982.

[25] An Hong-Zhi and Chen Zhao-Guo, On convergence of LAD estimates in autoregression with infinite varince. *Journal of Multivariate Analysis*, Vol. **12**, No. 3, p. 335—345, 1982.

[26] An Hong-Zhi, Chen Zhao-Guo and E. J. Hannan, Autocorrelation, autoregression and autoregressive approximation, *The Annals of Statistics*, Vol. **10**, No. 3, p. 926—936, 1982.

[27] B. G. Quinn and D. F. Nicholls, The estimation of random coefficint autogressive models. II, *Journal of Time Series Analysis*, Vol. **2**, No. 3, p. 185—203, 1981.

[28] E. J. Hannan, The estimation of the order of an ARMA process, *The Annals of Statistics*, Vol. **8**, No. 5, p. 1072—1081, 1980.

[29] E. J. Hannan, Dunsmuir, W. T. M. and Deistler, M. Estimation of vector ARMAX models, *J. Multivariate Anal.*, **10**, p. 275—295, 1980.

[30] E. J. Hannan and J. Rissanen, Recursive estimation of mixed autoregressive moving averageorder, *Biomatrica*, **69**, p. 81—94, 1982.

[31] D. A. Jones, Non-linear models in time series, *J. R. Statist. Soc. B*, Vol. **43**, No. 2, 1978.

[32] Tong, H. and K. S. Lim, Threshold autoregression, limit cycles and cyclical data, *J. of the RSS.*, Series B, **2**, No. 3, p. 245—292, 1980.

[33] 安鸿志、成平，自回归参数估计的渐近均方误差，应用数学学报，Vol. **3**, No. 1, 1980.

[34] 黄文杰、曹鸿兴、顾岚、项静恬，时间序列的 ARIMA 季节模型在长期预报中的应用，科学通报，Vol. **25**, No. 22, p. 1030—1033, 1980.

[35] 徐立子、杜金观、项静恬、顾岚，随机序列的新息方法在电力系统负荷预报中的应用，科学通报，Vol. **25**, No. 22, p. 1036—1039, 1980.

[36] 曹鸿兴、罗乔林，气象历史序列的最大熵谱分析，科学通报，第 8 期，1979.

[37] 郑大伟、赵铭，自回归谱技术用于天文和地球动力学，天文学报，Vol. **20**, No. 3, 1979.

[38] 郑大伟、黄慧玉、廖德春、罗时芳，用 AR 序列改进世界时服务，天文学报，Vol. **21**, No. 2, 1980.

后　记

本书成于 1976 年前后，它包括了近代时间序列分析中较活跃的一部分. 几年来，书中所介绍的方法和理论都有了新的发展，这些内容来不及反映在书中. 此外，书中也未收集 ARMA 模型的定阶方法、非线性模型及多维序列的统计方法，而这些内容目前也已在文献中丰富起来. 不过，本书的线性 ARMA 模型分析仍是最基本的，是了解时间序列分析中更新更广泛的内容的基础. 为了有兴趣的读者进一步掌握更多更新时序分析方面的知识，在出版前夕，我们增添了一些新的文献目录(从[23]至[38])，并且给出以下的分类，供读者查阅.

一、ARMA (p,q) 模型的定阶方法(见[23][24][28][30]).

二、非线性模型分析(见[31][32]).

三、多维 ARMA 模型的统计分析(见[15][29]).

四、方差为无穷的 AR 模型分析(见[25]).

五、系数为随机量的 AR 模型分析(见[27]).

六、ARMA 模型估计的渐近性质(见[16][26][33]).

七、应用及其它(见[34]至[38]).

作　者　1982.12.

后　记

本书成于1976年前后，它包括了近代时间序列分析中较活跃的一部分．几年来，书中所介绍的方法和理论都有了新的发展，这些内容来不及反映在书中．此外，书中也未收集ARMA模型的定阶方法、非线性模型及多维序列的统计方法，而这些内容目前也已在文献中丰富起来．不过，本书的线性ARMA模型分析仍是最基本的，是了解时间序列分析中更新更广泛的内容的基础．为了有兴趣的读者进一步掌握更多更新时序分析方面的知识，在出版前夕，我们增添了一些新的文献目录(从[23]至[38])，并且给出以下的分类，供读者查阅．

一、ARMA (p,q) 模型的定阶方法(见[23][24][28][30])．

二、非线性模型分析(见[31][32])．

三、多维 ARMA 模型的统计分析(见[15][29])．

四、方差为无穷的 AR 模型分析(见[25])．

五、系数为随机量的 AR 模型分析(见[27])．

六、ARMA 模型估计的渐近性质(见[16][26][33])．

七、应用及其它(见[34]至[38])．

作　者　1982.12.